NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AN ANALYSIS OF U.S. ARMY CADET COMMAND ADVANCED CAMP PERFORMANCE AND SURVEY DATA

by

Elizabeth A. Bochner

March 2019

Thesis Advisor: Robert A. Koyak
Second Reader: Matthew Norton

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2019	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
AN ANALYSIS OF U.S. ARMY CADET COMMAND ADVANCED CAMP PERFORMANCE AND SURVEY DATA

5. FUNDING NUMBERS

6. AUTHOR(S) Elizabeth A. Bochner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Cadet Command, Fort Knox, KY 40121

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

In this thesis, we use several statistical techniques to identify subsets of Reserve Officer Training Corps (ROTC) cadets who demonstrate a propensity for strong or weak performance at Advanced Camp, which is the final summer training event in the ROTC program designed to train and evaluate officer-candidates in various aspects of military performance. We analyze cadet survey responses prior to Advanced Camp to gain insights about levels of preparation for the rigorous training event. Finally, we develop models to predict a cadet's Advanced Camp performance score and his or her order of merit listing (OML), or rank among same-year ROTC cadets. A cadet's OML impacts his or her branch assignment upon commission. We use data on approximately 6,000 ROTC cadets from Advanced Camp in the summer of 2018 to conduct our analysis. We find that cadets who attend academically rigorous schools outperform their peers at Advanced Camp and in overall OML. Additionally, we find that cadets who receive two-year active duty scholarships or Green to Gold hip pocket scholarships also outperform cadets who receive alternative scholarships at Advanced Camp and in OML. Our predictive models successfully project a cadet's Advanced Camp performance score and OML ranking. Ultimately, we find that Advanced Camp performance scores are a crucial component of a cadet's OML.

14. SUBJECT TERMS Army, cadet, ROTC, Advanced Camp, survey, OML	**15. NUMBER OF PAGES** 91
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

AN ANALYSIS OF U.S. ARMY CADET COMMAND ADVANCED CAMP PERFORMANCE AND SURVEY DATA

Elizabeth A. Bochner
Lieutenant, United States Navy
BS, U.S. Naval Academy, 2013

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 2019

Approved by: Robert A. Koyak
Advisor

Matthew Norton
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

In this thesis, we use several statistical techniques to identify subsets of Reserve Officer Training Corps (ROTC) cadets who demonstrate a propensity for strong or weak performance at Advanced Camp, which is the final summer training event in the ROTC program designed to train and evaluate officer-candidates in various aspects of military performance. We analyze cadet survey responses prior to Advanced Camp to gain insights about levels of preparation for the rigorous training event. Finally, we develop models to predict a cadet's Advanced Camp performance score and his or her order of merit listing (OML), or rank among same-year ROTC cadets. A cadet's OML impacts his or her branch assignment upon commission. We use data on approximately 6,000 ROTC cadets from Advanced Camp in the summer of 2018 to conduct our analysis. We find that cadets who attend academically rigorous schools outperform their peers at Advanced Camp and in overall OML. Additionally, we find that cadets who receive two-year active duty scholarships or Green to Gold hip pocket scholarships also outperform cadets who receive alternative scholarships at Advanced Camp and in OML. Our predictive models successfully project a cadet's Advanced Camp performance score and OML ranking. Ultimately, we find that Advanced Camp performance scores are a crucial component of a cadet's OML.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AD	active duty
ADP	Army Doctrine Publication
APFT	Army physical fitness test
ARNG	Army National Guard
BRM	basic rifle marksmanship
C	capable
CBRN	chemical, biological, radiological, nuclear
CC	confidence course
CFF	call for fire
COER	Cadet Officer Evaluation Report
CV	cross-validation
CY	calendar year
E	excellent
EMPLID	employee identification
FY	fiscal year
GCV	generalized cross-validation
JROTC	Junior Reserve Officer Training Corps
MARS	multivariate adaptive regression spline
O	outstanding
OML	order of merit listing
OMS	outcome metrics score
P	proficient
PC	principal component
PCR	principal component regression
PLS	partial least squares
PT	physical training
RMSE	root-mean-square error
ROTC	Reserve Officer Training Corps
RSS	residual sum of squares
TC3	tactical combat casualty care

U	unsatisfactory
USACC	United States Army Cadet Command
USMC	United States Marine Corps

EXECUTIVE SUMMARY

Each year, approximately 6,000 Reserve Officer Training Corps (ROTC) cadets who are between their junior and senior years of college attend a rigorous summer training event called Advanced Camp. Advanced Camp, which is designed to train and evaluate officer-candidates in various aspects of military performance, emphasizes marksmanship, physical fitness, leadership, and squad-level exercises. Cadets receive scores for their performance at Advanced Camp. Advanced Camp scores are an important component of a cadet's order of merit listing (OML) which impacts his or her branch assignment upon commission. During the summer of 2018, cadets who attended Advanced Camp completed surveys on how effective their preparation for Advanced Camp was, based on their ROTC program training and personal performance metrics. Our research investigates over 6,000 observations of cadets who attended Advanced Camp in 2018. We aim to identify subsets of cadets that show a propensity for strong or weak performance at Advanced Camp. Cadet surveys conducted prior to attending Advanced Camp provide insights behind subset preparation. We also aim to develop an effective hierarchical model to predict a cadet's Advanced Camp performance score and ultimately his or her OML.

This thesis answers the following research questions:

1. How well are students prepared for Advanced Camp? In other words, which subsets of cadets, based on cadet data and demographics, show a propensity for strong or weak performance at Advanced Camp?

2. What insights can the surveys provide about preparation for Advanced Camp?

3. Can a hierarchical model effectively predict overall Advanced Camp performance scores and subsequently OML? How important is Advanced Camp for determining OML?

We identify several subsets of cadets that show a propensity for strong performance at Advanced Camp, and subsequently find themselves in the upper echelons of OML

ranking. They are cadets who attend institutions with difficult admissions standards, and cadets who receive either the two-year active duty scholarship or the Green-to-Gold hip pocket scholarship. Cadets in these subsets outperform their peers who attend institutions with less stringent admissions standards and who receive other types of scholarships.

We use a classification tree with a penalizing loss matrix to predict a cadet's Advanced Camp performance score within one degree of his or her actual score using strictly pre-existing cadet data and demographic information. We find the most significant predictor variable in this model to be the cadet's performance ranking from his or her junior year of ROTC, as this creates the first split that occurs in the classification tree.

We assess a variety of modeling techniques to predict a cadet's OML. We determine that a multivariate adaptive regression spline (MARS) model produces the most accurate predictions. We tune and use a MARS model to predict a cadet's OML both with and without Advanced Camp performance variables to compare the predictive power of each model and to observe the impact of a cadet's Advanced Camp performance on OML. A transformation of our predicted results ensures that the values are on the same scale as the original rankings. Table ES-1 shows both the untransformed and transformed root-mean-square errors (RMSE) for the different predictive modeling techniques we explore in our study with Advanced Camp performance variables included. We see that the MARS model has the lowest RMSE, making it the most accurate of the alternatives that we consider.

Table ES-1. RMSE for Predictive Models with Advanced Camp Variables

	Predictive Models with Advanced Camp Variables	
	RMSE: Untransformed OML Response	*RMSE: Transformed OML Response*
Multiple Regression	303.1	203.3
PCR	305.1	204.6
PLS	303.1	203.4
Elastic Net	303.0	203.1
Random Forest	274.9	257.9
MARS	227.8	192.0

Our transformed MARS model outperforms all of the alternatives we explore in our study. We find that the MARS model predicts a cadet's OML rank within 500 places of his or her actual rank 98.4% of the time, within 200 places of his or her actual rank 88.5% of the time, and within 100 places of his or her actual rank 76.9% of the time.

Table ES-2 shows both the untransformed and transformed (RMSE) for the different predictive modeling techniques, we explore in our study without the Advanced Camp performance variables.

Table ES-2. RMSE for Predictive Models without Advanced Camp Variables

	Predictive Models without Advanced Camp Variables	
	RMSE: Untransformed OML Response	*RMSE: Transformed OML Response*
Multiple Regression	457.4	402.8
PCR	459.5	403.8
PLS	457.4	402.9
Elastic Net	457.4	402.5
Random Forest	423.7	413.0
MARS	406.5	385.9

Again, the transformed MARS model outperforms all of the alternatives we explore in our study. We find that this MARS model predicts a cadet's OML rank within 500 places of his or her actual rank 91.9% of the time, within 200 places of his or her actual rank 74.0% of the time, and within 100 places of his or her actual rank 64.0% of the time.

We find that the absence of Advanced Camp performance variables in our predictive model for OML substantially reduces prediction accuracy, which demonstrates the importance of strong Advanced Camp performance.

Based on results from our study, we believe that investing in more cadets at academically rigorous institutions may elevate ROTC performance both at Advanced Camp and in overall OML. Additionally, re-allocating scholarships by increasing the number of two-year active duty scholarships and Green-to-Gold hip pocket scholarships may also enhance cadet performance throughout the ROTC program. We also believe that Advanced Camp performance is more valuable to OML than the ROTC program designates in their OML model based on little variance in the other contributing

components. To investigate what the true weight of Advanced Camp performance scores should be, we recommend a longitudinal study that tracks ROTC-produced Army officers throughout their initial active duty commitments to observe the correlation between a cadet's Advanced Camp performance score and his or her corresponding performance as an officer.

I. INTRODUCTION

A. MOTIVATION AND OBJECTIVES

During the summer between their junior and senior years of college, all U.S. Army Reserve Officer Training Corps (ROTC) cadets attend Advanced Camp as a part of their training program. Advanced Camp, which is designed to train and evaluate officer-candidates in various aspects of military performance, emphasizes marksmanship, physical fitness, leadership, and squad-level exercises. Cadets must complete Advanced Camp in order to earn their commissions. Advanced Camp performance is a significant component of a cadet's order of merit listing (OML), or rank among same-year ROTC cadets, which can impact his or her branch assignment upon commission. As of fiscal year (FY) 2018, approximately 6,000 cadets from around the country attend Advanced Camp every summer; they come from ROTC units located at colleges of varying sizes and academic ratings (Haupt 2018b). With multiple options for joining ROTC, cadets attending Advanced Camp will have enrolled in the ROTC program for varying lengths of time by the time they earn their undergraduate degrees. ROTC is the biggest producer of new Army officers: as of FY 2017, nearly 60% of commissioned officer gains in the Army came from ROTC programs (Office of the Under Secretary of Defense, Personnel, and Readiness 2017).

During the summer of 2018, cadets who attended Advanced Camp completed surveys on how effective their preparation for Advanced Camp was, based on their ROTC program training and personal performance metrics. The purpose of our study is to identify cadet subsets and corresponding survey responses that suggest a propensity for both strong and weak Advanced Camp performance scores. Ultimately, this research aims to build a hierarchical model to first predict a cadet's overall Advanced Camp performance score and then subsequently to predict the same cadet's OML.

Our research investigates over 6,000 observations of ROTC cadets in the FY 2018 Advanced Camp. The goal of our research is to inform the U.S. Army Cadet Command (USACC) of the results of our analysis in order to improve the overall quality of the U.S.

Army officer corps. Additionally, our research suggests areas where the USACC can make future investments, specifically regarding ROTC scholarship allocation.

B. FOCUS OF THE RESEARCH

It is in the mission of the USACC to ensure that cadets are adequately prepared for all graded events at Advanced Camp to produce capable Army officers upon graduation and commissioning. To achieve this objective, the USACC seeks to identify which subsets of cadets demonstrate a propensity for strong or weak performance scores at Advanced Camp. In particular, the USACC wants to determine if survey responses suggest any prior indicators for the strong or weak performance scores. Survey insights may contain valuable lessons that can enhance the ROTC training program in years to come. Our study focuses on developing an analytical approach to investigating factors influencing Advanced Camp performance in order to build a hierarchical predictive model for Advanced Camp performance and OML.

Our research answers the following questions:

1. How well are students prepared for Advanced Camp? In other words, which subsets of cadets, based on cadet data and demographics, show a propensity for strong or weak performance at Advanced Camp?

2. What insights can the surveys provide about preparation for Advanced Camp?

3. Can a hierarchical model effectively predict overall Advanced Camp performance scores and subsequently OML? How important is Advanced Camp for determining OML?

C. ORGANIZATION OF THIS THESIS

In this chapter, Chapter I, we described the motivation, objectives, research focus, and structure of this thesis.

In Chapter II, we discuss in greater detail the mission of USACC, the structure of the Advanced Camp summer training event, and the 2018 Advanced Camp survey. We also provide a review of prior work that is related to the focus of our research.

In Chapter III, we describe our analysis of the cadet data and survey results in relation to Advanced Camp performance, and we describe the hierarchical predictive modeling approach that we adopt. This chapter also describes the statistical methods we use for each section.

In Chapter IV, we present our results for both the analysis of the cadet data and survey results and the hierarchical predictive modeling. Additionally, we discuss the strengths and weaknesses of the processes and models surrounding the results.

In Chapter V, we state our conclusions and propose recommendations for future work.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. USACC

The U.S. Army established the Cadet Command at Fort Monroe, Virginia, in 1986 to centralize the Reserve Officer Training Corps (ROTC) and Junior ROTC (JROTC) programs (U.S. Army Cadet Command 2017c). The USACC currently is headquartered at Fort Knox, Kentucky, and oversees over 20,000 ROTC cadets in 273 ROTC programs across the United States. According to the mission statement, the USACC "partners with universities to recruit, educate, develop, and inspire Senior ROTC cadets to commission officers of character for the Total Army; and partners with high schools to conduct JROTC to develop citizens of character for a lifetime of commitment and service to the nation" (U.S. Army Cadet Command 2017c). The USACC is organized into eight brigades, each of which is responsible for the ROTC programs in its region of the country. A display of the brigade geographic distribution is shown in Figure 1.

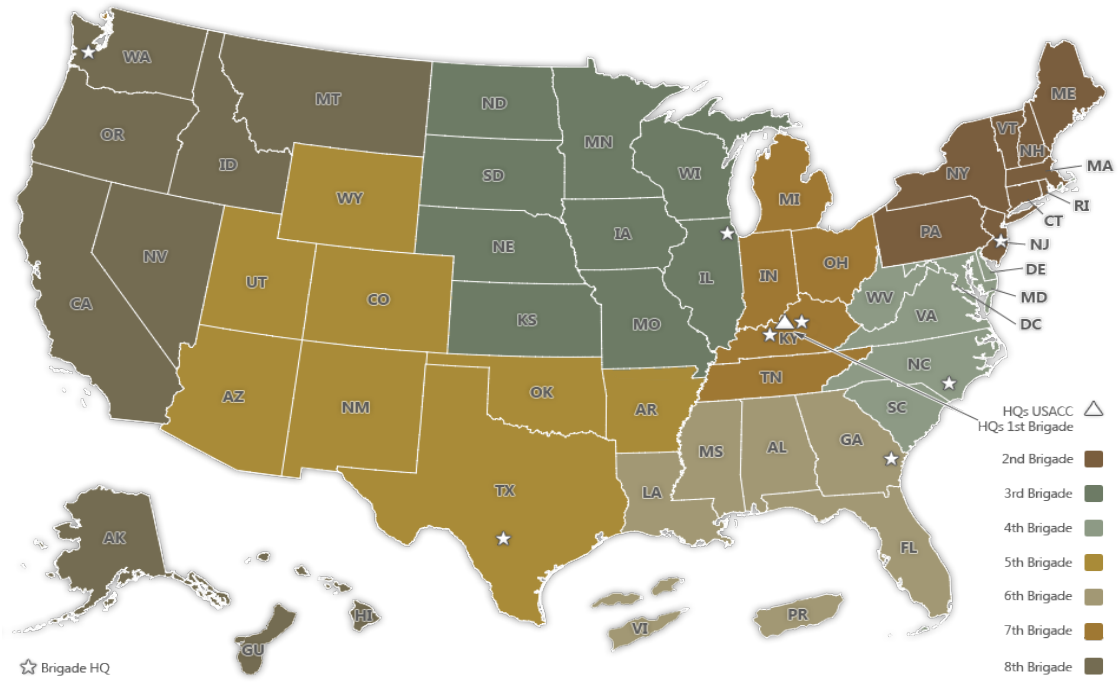

Figure 1. USACC Brigade Geographic Distributions. Source: U.S. Army Cadet Command (2017b).

Every year, USACC has a requirement, or mission, to produce a specific number of Second Lieutenants through the ROTC program. There are several options for students to join the ROTC program to help meet this mission. A high-school student can earn a four-year ROTC scholarship through a national scholarship process. Additionally, students may enroll in the ROTC program upon arrival at their undergraduate institutions. Students who enroll in ROTC without a scholarship can earn a three- or two-year scholarship during their first or second years of school, respectively. Finally, if students decide at the end of their sophomore year that they want to join the ROTC program, they may do so upon completion of Basic Camp. Basic Camp is a program offered during the summer between a student's sophomore and junior years to teach new cadets about the Army. Figure 2 illustrates the options for joining ROTC.

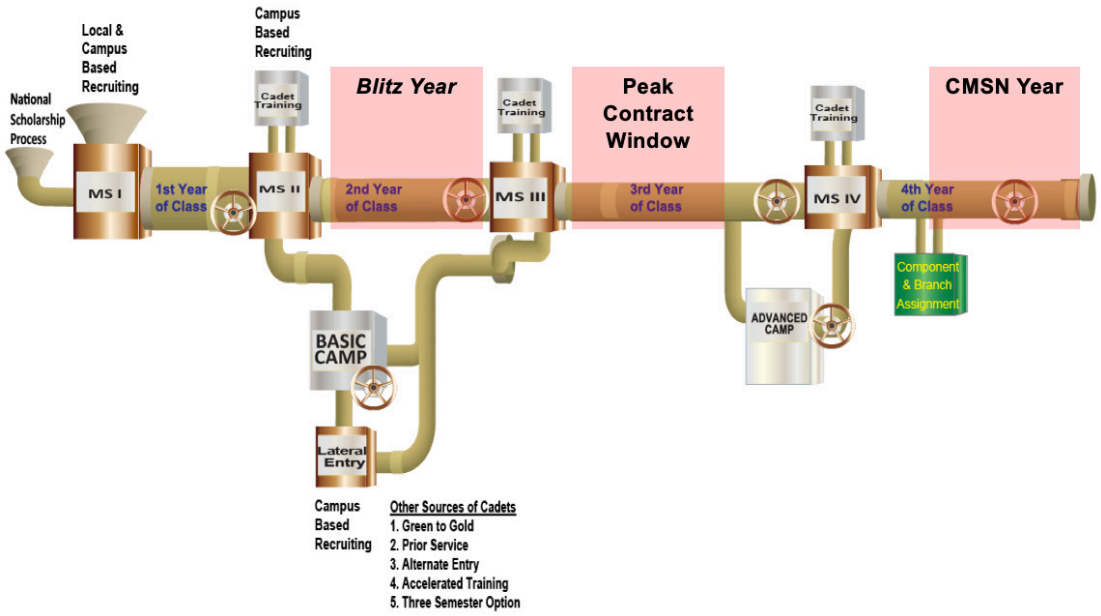

Figure 2. ROTC Development Model. Source: Haupt (2018b).

B. ADVANCED CAMP

Advanced Camp is a rigorous 31-day summer training event that takes place in Fort Knox, Kentucky. The mission of Advanced Camp is to "train U.S. Army ROTC cadets to Army standards and to develop leadership and evaluate officer potential" (U.S. Army Cadet

Command 2017a). Advanced Camp consists of five phases. Phase I focuses on physical training, rules of engagement, and background knowledge for future phases. In Phase II, cadets begin team building through physical exercises, the Field Leaders' Reaction Course, land navigation, and marksmanship. Phase III transitions into platoon-level field exercises. Cadets return from the field and begin Phase IV, which involves cleaning gear and recovering from the field. In Phase V, cadets receive Cadet Officer Evaluation Reports (COER) and review their progress over the course of Advanced Camp. Graded events take place throughout all five phases resulting in a cumulative Performance Summary score for each cadet. Table 1 lists the graded events at Advanced Camp; these graded events are of particular interest to this study as potential predictors of outcomes such as OML.

Table 1. Graded Events at Advanced Camp. Adapted from Duncombe (2018).

Physical Fitness	
Event & Metric	**Possible Scores**
Foot March: Status	< 3 Hours; Go; No Go
Foot March: Time	Hours:Minutes
Army Physical Fitness Test Score	Scores range from 0 - 300
Marksmanship	
Event & Metric	**Possible Scores**
Call for Fire (CFF): Status	First Time Go; Go; No Go
Alternate Course: Qualification	Unqualified, Marksman, Sharpshooter, Expert
Alternate Course: Attempts	Attempts range from 0 - 12
Alternate Course: Score	Scores range from 0 - 40
Land Navigation	
Event & Metric	**Possible Scores**
Written Exam	Score from 0 - 100
Practical Exercise: Status	First Time Go; Go; No Go
Practical Exercise: # Targets Found	Scores range from 0 - 6
Pop-Up Range: Qualification	Unqualified, Marksman, Sharpshooter, Expert
Pop-Up Range: Score	Scores range from 0 - 40
Leadership	
Event & Metric	**Possible Scores**
Garrison 1, Garrison 2, Field 1, and Field 2 Positions	Unsatisfactory, Capable, Proficient, Excellent
Cadet Rank & Cadet Total	Rank in Platoon/Size of Platoon
Other	
Event & Metric	**Possible Scores**
Tactical Combat Casualty Care (TC3)	First Time Go; Go; No Go
Chemical, Biological, Radiological, Nuclear (CBRN)	First Time Go; Go; No Go
Confidence Course (CC)	First Time Go; Go; No Go
Overall Performance Summary Score	
O: Outstanding E: Excellent P: Proficient C: Capable U: Unsatisfactory	

C. ADVANCED CAMP SURVEY

During the summer of 2018, cadets who attended Advanced Camp completed a survey on the effectiveness of their preparation for Advanced Camp, based on their ROTC program training and personal performance metrics. Although the USACC does not require cadets to complete the survey, the data provided from it may provide insight into factors that affect Advanced Camp performance. Cadets took the survey before arriving at Advanced Camp. The survey has a 46% response rate, and although this is not a full response, the data may still provide insights about factors that impact Advanced Camp performance.

The survey is administered as a computer-based questionnaire. Most questions on the survey elicit responses on five- or seven-point Likert scales. Several questions call for short-answer responses. The survey also contains several yes-or-no questions. The questions used in the survey are shown in Appendix A.

D. THE ORDER OF MERIT LIST (OML) MODEL

The Order of Merit List (OML) model is designed to translate a cadet's ROTC achievements, academic standing, and extracurricular involvement into an objective ranking within his or her class year. The OML model produces an Outcome Metrics Score (OMS) for each cadet. Cadets are ranked from 1 to n based on their OMS, where n is the total number of cadets in the ROTC class year. A rank of 1 indicates the top performance in the ROTC class year; a rank of n represents the lowest performance. For the FY 2019 cohort, the composition of the OMS that is used to derive the OML model is shown in Figure 3.

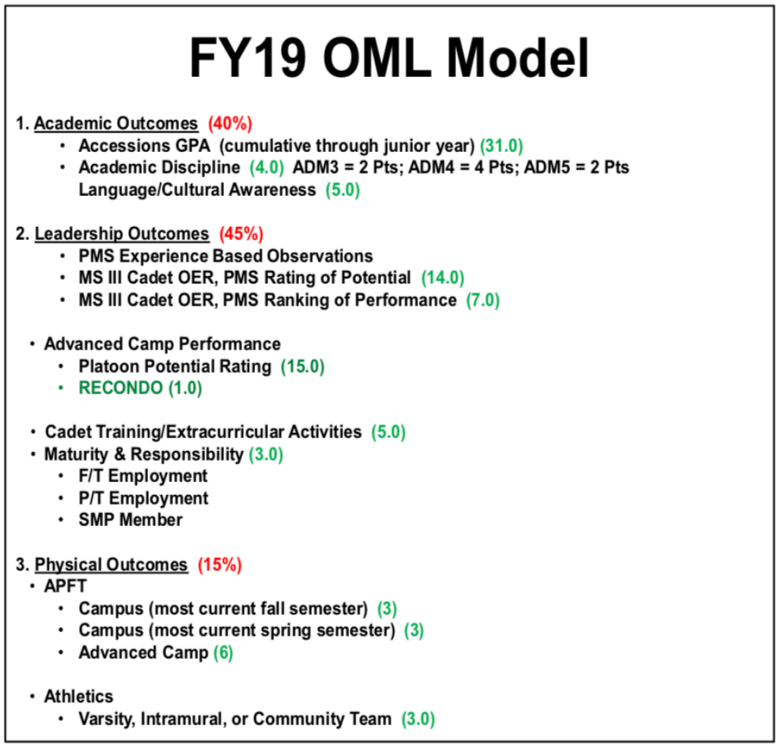

Figure 3. FY 2019 OML Model. Source: U.S. Army Cadet Command (2018b).

A cadet's OML is a significant factor in his or her commissioning status as either active duty (AD) or Army National Guard (ARNG). Additionally, OML impacts a cadet's branch assignment, or occupational specialty, upon commissioning. The higher a cadet is on the list, the more likely he or she is to receive his or her first-choice assignment.

E. LITERATURE REVIEW

To our knowledge, no previously published studies have examined the questions posed in our thesis with respect to ROTC cadets. There are, however, several prior publications that bear on topics that are similar to those that we examine. The first document is an article from July 2017 written by then-commanding general of the U.S. Army Cadet Command, Maj. Gen. Christopher P. Hughes. The second document is a Naval Postgraduate School thesis from 2016 by Ben McCaleb III. In this section, we discuss these publications in relation to our research.

Hughes (2017) references Army Doctrine Publication (ADP) 6–22: Army Leadership, which outlines the principles and requirements of U.S. Army leaders (Department of the Army 2012b). Hughes argues that a rigorous summer training program is essential for the ROTC program to produce second lieutenants that meet the requirements of ADP 6–22. Hughes describes the process of ensuring that cadre, or instructors, receive consistent training to better support the cadets at Advanced Camp. He suggests that well-trained cadre provides the necessary foundation for increasing training intensity at Advanced Camp, which in turn ensures that ROTC will produce strong officers that meet the requirements of ADP 6–22.

McCaleb (2016) investigates U.S. Marine Corps (USMC) recruit characteristics that translate to successful careers into various occupational fields. Specifically, he focuses on whether the information available prior to a recruit enlisting can help the USMC predict the best occupational field for the recruit. To address this, he uses a multinomial regression with Elastic Net variable selection to predict the probability of a recruit achieving a particular Computer Tier Score Classification; the Computer Tier Score Classification is an ordinal response variable that represents a recruit's performance and qualification for re-enlistment. As in our study, the predictor variables in McCaleb's analysis are characteristics of recruits before their initial enlistment.

Both Hughes (2017) and McCaleb (2016) share similar motivations with our thesis. Hughes emphasizes the significance of a thorough and rigorous Advanced Camp to produce qualified officers in the Army. Understanding the demographic factors that contribute to strong and weak performances at Advanced Camp will help the USACC achieve Hughes' summer training goals. McCaleb's research highlights the explanatory power of pre-existing personnel characteristics for predicting performance metrics in the USMC. As in McCaleb's research, we use cadet demographics and performance data prior to Advanced Camp to predict the ordinal Advanced Camp performance response variable. Our research closely parallels the concepts in these documents.

III. DATA AND METHODOLOGY

A. DATA AND FORMATTING

In this section, we outline the data sets available for this study. Additionally, we provide a detailed summary of our data processing steps for this analysis.

1. Data Files

USACC provided all of the data for this study along with a data dictionary, a variable list, and the survey instrument in order to assist with our analyses. Specifically, our research focuses on two data sets. The first data set contains cadet data, demographic information, and survey responses. The information in this data set is specific to cadets who attended Advanced Camp in the summer of 2018. This information is specific to 5,087 cadets who are between their junior and senior years of college, of whom 2,377 cadets completed the survey.

The second data set contains OML-specific information on 23,496 cadets covering four years of ROTC, up to and including the cadets graduating in calendar year (CY) 2019. This data set includes 5,377 cadets who are between their junior and senior years of college.

2. Data Processing

The first step in our data-handling process is preparing the data. The first data set (cadet data, demographics, and survey responses) contains 10 duplicate entries based on the employee identification column (EMPLID) while the second data set (OML data) contains 208 duplicate entries. We remove these duplicate entries. Next, we merge the two data sets on the common EMPLID field and we remove any records that are not present in both data sets. Additionally, we remove any observations where the OML is 0 because this value is not a feasible OML rank. We convert all missing values to NA for consistency across the data set. We also remove three numerical columns that exhibit no variability (i.e., all values are the same). The merged data set has 4,814 observations on 121 variables; we lose 273 observations from our data processing.

B. VARIABLES USED IN THE STUDY

In this section, we describe the response and predictor variables we consider for use in our study.

1. Response Variables

Two response variables are of interest to meeting the objectives of our research: Advanced Camp performance scores and the OML ranking. The Advanced Camp performance score is a classification of a cadet's overall performance at Advanced Camp, accounting for all facets of the training including physical fitness, leadership, marksmanship, and land navigation. The possible scores are as follows, in order of best to worst performance: Outstanding, Excellent, Proficient, Capable, and Unsatisfactory. Definitions of these performance classifications are provided in Table 2. As discussed in Chapter 2, the OML ranking is a discrete variable that represents a cadet's overall ROTC standing among his or her peers.

Table 2. Advanced Camp Performance Definitions. Source: Duncombe (2018).

Advanced Camp Performance Scores	Meaning
Outstanding (O)	Top 15th Percentile
Excellent (E)	16-49th Percentile
Proficient (P)	50-85th Percentile
Capable (C)	Bottom 15th Percentile
Unsatisfactory (U)	Cadet Failed to Meet AC Course Standards

2. Predictor Variables

Our merged data set has 118 possible predictor variables which we organize into the following categories for ease of explanation: cadet data and demographics, Advanced Camp performance metrics, pre-survey questions, and OML model variables.

a. Cadet Data and Demographic Variables

The cadet data and demographic predictor variables in our data set are listed in Table 3. We also discuss several of these variables in the following text.

Table 3. Cadet Data and Demographic Predictor Variables. Adapted from U.S. Army Cadet Command (2018a).

Variable	Description	Type
W_GENDER	Gender	Categorical
W_REDCAT	Race/Ethnicity	Categorical
W_ENRL_STAT_CD	ROTC Enrollment Status	Categorical
W_CIP_CODE	Academic CIP Code	Categorical
W_ACD_DISC_MIX	Academic Discipline Mix	Categorical
W_SCHL_CAT_CD	ROTC Scholarship Category	Categorical
Brigade	ROTC Brigade	Categorical
Height	Cadet Height in Inches	Numeric
Weight	Cadet Weight in Pounds	Numeric
Host.Tier	ROTC Host Unit Academic Tier	Ordinal
Academic.Tier	ROTC Academic Tier	Ordinal

The W_ENRL_STAT_CD variable represents ROTC enrollment status, which indicates whether or not a cadet has an official ROTC contract and has been awarded a scholarship, or if the cadet is pending enrollment, disenrollment, or scholarships. The possible values of W_ENRL_STAT_CD are shown in Table 4.

Table 4. ROTC Enrollment Status Variable Definitions. Adapted from U.S. Army Cadet Command (2018a).

W_ENRL_STAT_CD Value	Description
C	Completion - All ROTC Requirements Met
E	Enrolled and Contracted
E$	Enrolled and Contracted, Scholarship
F	LOA - Pending Disenrollment
L	LOA - Pending Return
M	Commissioned
R	Scholarship Award Pending
X	Green to Gold Active Duty Option

The W_CIP_CODE variable describes a cadet's specific academic program. This categorical variable has over 300 levels; for analytical purposes we reduce the number of levels in this field by collapsing low-frequency values (those that contain less than 0.6% of cadets in the sample) into an 'Other' level. This reduction results in a W_CIP_CODE

variable having 32 levels. The W_ACD_DISC_MIX variable further groups the W_CIP_CODE programs into the five categories listed in Table 5.

Table 5. Academic Discipline Mix Categories. Adapted from U.S. Army Cadet Command (2018a).

W_ACD_DISC_MIX Value	Description
1	Generalist
2	Technical Management
3	Physical Science/Analytical
4	Engineering
5	Nursing

The W_SCHL_CAT_CD variable describes a cadet's scholarship status. A high-school student can earn a four-year ROTC scholarship through a national scholarship program. Additionally, students can enroll in the ROTC program upon arrival at their undergraduate institution. Students who enroll in ROTC without a scholarship can earn a three- or two-year scholarship during their first or second years of school, respectively. Half-year increments of scholarships are also available. Table 6 lists the possible values for the W_SCHL_CAT_CD variable, along with an explanation of each scholarship type.

Table 6. W_SCHL_CAT_CD Variable Descriptions. Adapted from U.S. Army Cadet Command (2018a).

W_SCHL_CAT_CD Values	Description
2C	2-Year Enrolled Competition
2H	2.5-Year Scholarship
3C	3-Year Enrolled Competition
3D	3-Year Advanced Designee
3H	3.5-Year Scholarship
4R	4-Year Regular Cycle
4U	4-Year HQCC Scholarship
A2	2-Year Active Duty
BC	Basic-Camp Active Duty
HP	2-Year Green-to-Gold Hip Pocket
N2	2-Year Non-Enrolled Competition
N3	3-Year Non-Enrolled Competition

The Academic.Tier variable describes the academic difficulty of a cadet's undergraduate institution where he or she attends classes, and the Host.Tier variable describes the academic difficulty of the undergraduate institution where a cadet attends ROTC. Usually, a cadet's Academic.Tier and the Host.Tier represent the same institution. However, in regions dense with colleges and universities, it is possible that one ROTC unit covers the entire region. In those instances, a cadet might attend ROTC at a different university from where he or she attends classes. The tier categories are based on the 2018 Peterson's Guide entrance difficulties for universities and are shown in Table 7.

Table 7. 2018 Peterson's Guide Tier Categories. Source: Haupt (2018a).

Tiers	Description
1	*Most Difficult* - More than 75% of all freshmen were in the top 10% of their high school class and scores over 1310 on SAT (critical reading and math scores combined) or over 29 on ACT; about 30% or fewer of all applicants accepted.
2	*Very Difficult* - More than 50% of all freshmen were in the top 10% of their high school class and scored over 1230 on SAT (critical reading and math scores combined) or over 26 on ACT; about 60% or fewer of applicants accepted.
3	*Moderately Difficult* - More than 75% of freshmen were in the top 50% of their high school class and scored over 1010 on SAT (critical reading and math scores combined) or over 18 on the ACT; about 85% or fewer of applicants accepted.
4	*Minimally Difficult* - Most freshmen were not in the top 50% of their high school class and scored somewhat below 1010 on SAT (critical reading and math scores combined) or below 19 on ACT; up to 95% of applicants accepted (not 100%).
5	*Noncompetitive* - Virtually all applicants accepted regardless of high school rank or test scores.
Not Reported	*Reporting requirements are voluntary and many schools do not report for various reasons.*

b. Advanced Camp Performance Variables

The Advanced Camp performance variables in our data set are shown in Table 8. We discuss several of these variables in this section.

Table 8. Advanced Camp Performance Variables. Adapted from U.S. Army Cadet Command (2018a).

Variable	Description	Type
G1_POSN	First Garrison Rated Position	Categorical
G1_RTNG	First Garrison Rating	Categorical
G2_POSN	Second Garrison Rated Position	Categorical
G2_RTNG	Second Garrison Rating	Categorical
F1_POSN	First Field Rated Position	Categorical
F1_RTNG	First Field Rating	Categorical
F2_POSN	Second Garrison Rated Position	Categorical
F2_RTNG	Second Garrison Rating	Categorical
ALTC	Alt C Qualification Rating	Categorical
POPUP	Pop Up Qualification Rating	Categorical
WRTN_SCORE	Land Navigation Written Test Score (Percentage)	Numeric
TargetsFound	# of Day Land Navigation Points Found During PE	Numeric
ALTC_Attempts	# of Alt C Attempts	Numeric
AltCScore	Alt C Range Score	Numeric
POPUP_Attempts	Number of Pop Up Attempts	Numeric
PopScore	Pop Up Range Score	Numeric
FootmarchTime	Foot March Time in Minutes	Numeric
Score	APFT Score	Numeric

The G1_POSN and G2_POSN variables represent the garrison positions that cadets may hold during Advanced Camp. The positions include platoon leader, platoon sergeant, and squad leader. Cadets hold two of these leadership roles over the course of Advanced Camp and receive scores for their performances on a scale of outstanding (O), excellent (E), proficient (P), or capable (C). The G1_RTNG and G2_RTNG variables use this scoring for the G1_POSN, and G2_POSN variables respectively. Similarly, the F1_POSN, F2_POSN, F1_RTNG, and F2_RTNG variables follow the same construct for the field positions that cadets may hold during Advanced Camp.

The ALTC, ALTC_Attempts, and ALTCScore variables give a cadet's scores for the Alternate Course marksmanship qualification metrics at Advanced Camp. The categorical ALTC variables are scored, from best to worst, as follows: expert, sharpshooter, marksman, and unqualified. The numeric ALTC_Attempts variable tells us how many Alternate Course attempts it takes for a cadet to pass. Finally, the ALTC_Score variable

has the cadet's numeric shooting score for the Alternate Course qualification. Similarly, the POPUP, POPUP_Attempts, and PopScore variables follow the same structure for the Popup Range marksmanship qualification.

Cadets receive scores for two physical fitness events at Advanced Camp: the Army Physical Fitness Test (APFT) and the 12-mile foot march. The Score variable represents a cadet's performance on the APFT. The components of the APFT includes pushups, sit-ups, and a two-mile run; scores from each component are added together to get the overall APFT score. Figure 4 displays the scoring system for each APFT component.

Army Physical Fitness Test Score Table

MALE PUSHUPS	
Repetitions	Points
71	100
70	99
69	97
68	96
67	94
66	93
65	92
64	90
63	89
62	88
61	86
60	85
59	83
58	82
57	81
56	79
55	78
54	77
53	75
52	74
51	72
50	71
49	70
48	68
47	67
46	66
45	64
44	63
43	61
42	60

FEMALE PUSHUPS	
Repetitions	Points
48	
47	
46	
45	
44	
43	
42	100
41	98
40	97
39	95
38	93
37	91
36	90
35	88
34	86
33	84
32	83
31	81
30	79
29	77
28	76
27	74
26	72
25	70
24	69
23	67
22	65
21	63
20	62
19	60

MALE/FEMALE SITUPS	
Repetitions	Points
82	
81	
80	
79	
78	100
77	98
76	97
75	95
74	94
73	92
72	90
71	89
70	88
69	87
68	84
67	82
66	81
65	79
64	78
63	76
62	74
61	73
60	71
59	70
58	68
57	66
56	65
55	63
54	62
53	60

MALE 2-MILE RUN	
Run Time	Points
12:36	
12:42	
12:48	
12:54	
13:00	100
13:06	99
13:12	97
13:18	96
13:24	94
13:30	93
13:36	92
13:42	90
13:48	89
13:54	88
14:00	86
14:06	85
14:12	83
14:18	82
14:24	81
14:30	79
14:36	78
14:42	77
14:48	75
14:54	74
15:00	72
15:06	71
15:12	70
15:18	68
15:24	67
15:30	66
15:36	64
15:42	63
15:48	61
15:54	60

FEMALE 2-MILE RUN	
Run Time	Points
15:36	100
15:42	99
15:48	98
15:54	96
16:00	95
16:06	94
16:12	93
16:18	92
16:24	90
16:30	89
16:36	88
16:42	87
16:48	85
16:54	84
17:00	83
17:06	82
17:12	81
17:18	79
17:24	78
17:30	77
17:36	76
17:42	75
17:48	73
17:54	72
18:00	71
18:06	70
18:12	68
18:18	67
18:24	66
18:30	65
18:36	64
18:42	62
18:48	61
18:54	60

Extract from DA Form 705 for the 17-21 age group.
Reference FM 7-22 Army Physical Readiness Training (OCT 12).
Run time is minutes and seconds.

Figure 4. APFT Score Table. Source: Department of the Army (2012a).

The FootmarchTime variable gives the number of minutes that it takes a cadet to complete the 12-mile foot march at Advanced Camp. Cadets carry a 35-lb rucksack while they complete the foot march.

c. Advanced Camp Survey Variables

Our merged data set contains 69 variables that represent the questions on the Advanced Camp survey. We remove several variables that do not pertain to performance such as the travel process to arrive at Advanced Camp and paperwork and pay-related concerns.

d. OML Model Variables

The OML model predictor variables in our data set are listed in Table 9. We describe several of these variables in this section.

Table 9. OML Model Predictor Variables. Adapted from U.S. Army Cadet Command (2018a).

Variable	Description	Type
W_ADM_OMS_SCORE	Academic Discipline Mix	Categorical
W_APF_CST_SCORE	Advanced Camp APFT Score	Numeric
W_APF_FAL_RAW_SCOR	MS III Fall APFT Score	Numeric
W_APF_SPR_RAW_SCOR	MS III Spring APFT Score	Numeric
W_ACCESSION_GPA	Academic & ROTC GPA (Cumulative through Junior Year)	Numeric
W_ACTIV_PTS_EARNED	Athletic Points (Varsity, Intramural, Community)	Numeric
W_ATHS_OMS_SCORE	Athletics Score	Numeric
W_CST_RNK_SCORE	Advanced Camp Ranking of Performance	Numeric
W_CST_OER_SCORE	Advanced Camp Rating of Potential	Numeric
W_CLA_SCORE	Language & Cultural Awareness Score	Numeric
W_MS3_OER_SCORE	MS III Rating of Potential	Numeric
W_MS3_RNK_SCORE	MS III Ranking of Performance	Numeric
W_EACTI_PTS_EARNED	Cadet Training & Extracurricular Activity Points	Numeric
W_TRN_PTS_EARNED	Cadet Training Points Earned	Numeric
W_TRN_EXTRACUR_PTS	Cadet Extracurricular Points Earned	Numeric
W_TRN_ACT_SCORE	Cadet Activity Score	Numeric
W_MAT_OMS_SCORE	Maturity & Responsibility Points	Numeric

The W_MS3_OER_SCORE and W_MS3_RNK_SCORE variables give information about a cadet's leadership ability based on their junior year in ROTC. The W_MS3_OER_SCORE can take on the values 0, 25, 50, 75, or 100. The W_MS3_RNK_SCORE can take on any integer value from 0 to 100. Similarly, the

W_CST_OER_SCORE and the W_CST_RNK_SCORE variables give information about a cadet's leadership ability based on their Advanced Camp performance using the same values.

The W_ATHS_OMS_SCORE variable measures a cadet's involvement in various levels of athletics. The scoring systems for this variable is shown in Table 10.

Table 10. Athletics Scoring for OMS. Source: Duncombe (2018).

Athletics	Point Per Year	Max Points
Varsity	10	30
Intramural	5	15
Community	5	15
Total Possible Points		60

The USACC Circular 601–19-1 from June 2018 provides amplifying information about the OML variables we include in our data set (U.S. Army Cadet Command 2018b).

C. METHODOLOGY

In this section, we explain our approach for our hierarchical modeling of Advanced Camp performance and OML ranking. We then discuss the statistical methods we use to conduct our analysis of the cadet data and survey results in relation to Advanced Camp performance.

1. Classification Tree for Advanced Camp Performance

Our study uses a classification tree with a loss matrix to predict Advanced Camp performance based on pre-existing cadet data (Faraway 2016, p. 354). Classification trees are effective tools for multinomial response variables. In our study, the Advanced Camp performance variable is ordinal. The first node of a classification tree represents the most significant predictor variable. Nodes are split such that the classes present in a split node are predominantly of one class. The purity of the nodes in our classification tree is measured using the Gini index shown in Equation 1, where D_i represents the purity of node i, and p_{ik} represents the observed proportion of class k in node i (Faraway 2016, p. 354).

$$\boldsymbol{D_i = 1 - \sum_k p_{ik}^2} \quad (1)$$

Classification trees work to maximize the purity of a node and decrease its deviance. We incorporate a loss matrix into our classification tree to create penalties for misclassifications, reflecting the fact that the Advanced Camp rating is an ordinal variable. The loss matrix for our model is shown in Equation 2.

Actual

Predicted	**O**	**E**	**P**	**C**	**U**
O	**0**	**1**	**2**	**3**	**4**
E	**1**	**0**	**1**	**2**	**3**
P	**2**	**1**	**0**	**1**	**2**
C	**3**	**2**	**1**	**0**	**1**
U	**4**	**3**	**2**	**1**	**0**

(2)

Our loss matrix penalizes the classification tree for every degree of misclassification it gives a cadet. For example, if a cadet earns the best possible score, O, but the tree classifies the cadet as P, the model will incur a loss of 2. A generalized formula for the Gini index when incorporating a loss matrix into a classification tree is shown in Equation 3 where $L(i,j)$ indicates the loss associated with assigning class j to an observation with actual class i, and p_i and p_j represent the probabilities of classes i and j respectively (Therneau and Atkinson 2018, p. 7).

$$\boldsymbol{G(p) = \sum_i \sum_j L(i,j)\, p_i\, p_j} \quad (3)$$

We produce a confusion matrix by validating our model with a test set, which is a set of data that is not used in model estimation. A confusion matrix is a table that summarizes the performance of our model by displaying predicted classifications along the rows and the actual classifications along the columns. The diagonal values of the confusion matrix represent accurate predictions. We point-wise multiply the confusion matrix and the loss matrix, and then sum, to obtain an overall value for loss from our classification tree model. To make our results interpretable, we divide the overall loss value by the number of samples in the model to get average loss. Average loss represents our model's measure of performance for our study.

2. Multivariate Adaptive Regression Spline (MARS) Models for OML Prediction

A MARS model combines elements of linear regressions and step functions to capture non-linear characteristics of predictor variables (Hastie et al. 2017). We provide some background on these elements and discuss the applicability of MARS for our study.

a. Linear Regressions

Linear regression is a widely used technique for combining a collection of k predictor variables to produce a prediction for an outcome variable, Y. The general formula for a linear regression model is shown in Equation 4 (Devore 2016).

$$\boldsymbol{Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon} \quad (4)$$

In Equation 4, the β_0 term is the y-intercept of the model. The coefficients $\beta = \{\beta_1, \beta_2, \ldots, \beta_k\}$ represent the coefficients of the predictor variables. The variable ε is a random error term that is assumed to be normally distributed with expected value, $E(\varepsilon) = 0$, and variance, $V(\varepsilon) = \sigma^2$. The linear regression model assumes that the relationship between the predictor variables and the outcome variable can be described by a linear function with unknown parameters, plus a random error term. If the underlying relationship is nonlinear, the accuracy of the model may be negatively impacted. One way to incorporate nonlinear relationships within a linear regression model is to include polynomial parameters. An example of a polynomial regression model using a single predictor variable (x) is shown in Equation 5 (Devore 2016).

$$\boldsymbol{Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_d x^d + \varepsilon} \quad (5)$$

b. Step Functions

Step functions are another alternative for modeling nonlinear relationships. Rather than using nonlinear functions that are applied everywhere, step functions break the predictor variable into separate bins and fit different coefficients for each bin. A formula for binned linear regression with a single predictor variable is shown in Equation 6 (Boehmke 2018).

$$\boldsymbol{Y = \beta_0 + \beta_1 C_1(x) + \beta_2 C_2(x) + \cdots + \beta_k C_k(x) + \varepsilon} \quad (6)$$

In Equation 6, $C_j(x)$ is an indicator function that is equal to 1 if x lies in the interval $[a_{j-1}, a_j)$ and 0 otherwise, for all $j = 1, \ldots, k$.

c. MARS Models

Polynomial regression and step functions are effective predictive tools; however, both techniques require the user to manually specify how to use them, such as the degree of a polynomial or how to bin specific predictor variables. MARS models "capture the non-linearity aspect of polynomial regression by assessing cut points (knots) similar to step functions" (Hastie et al. 2017). The MARS approach involves searching across a predictor variable's values for a point at which two different linear relationships exist between the predictor and response variables. Multiple linear relationships may be appropriate for a given predictor variable. Places where a linear relationship shifts are called knots. MARS models may be pruned at knots that are not strong predictors.

We use a MARS model that prunes knots based on an expected change in R^2 of less than 0.001 through a generalized cross-validation (GCV) procedure (Boehmke 2018). For our study, we develop a MARS model that incorporates all of the OML model variables, which includes the Advanced Camp-specific variables. We also create a second MARS model that includes all of the OML model variables with the exception of the Advanced Camp-specific variables. We compare the model outcomes in order to evaluate the predictive power of our models when Advanced Camp performance is both known and unknown.

3. Useful Techniques

We use two nonparametric statistical techniques to explore cadet data and survey variables as they relate to Advanced Camp performance. They are the Jonckheere-Terpstra test for ordered differences and the Kruskal-Wallis test for differences in mean ranks among groups. The Kruskal-Wallis test is a one-way analysis of variance based on ranks where the grouping variable is categorical. The Jonckheere-Terpstra test is similar to the Kruskal-Wallis test, but applied in situations where the grouping variable is ordered.

Further discussion of these two widely used statistical techniques can be found in Kloke and McKean (2014).

4. Model Validation

Prior to conducting our analyses, we partition our data into a "training" set and a "test" set to properly assess the predictive power of our models. We randomly select 80% of our observations for the training set and 20% of our observations for the test set. We train our classification tree and MARS models with the training set, and then apply those models to the test set to assess their accuracy. Comparing the predicted observations to the actual observations allows us to determine the effectiveness of our models.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS

In this chapter, we present the results from the statistical tests and models we described in Chapter III. We provide the results of cadet data and survey analyses followed by the results of our predictive modeling for Advanced Camp performance and OML.

A. DEMOGRAPHIC DATA AND SURVEY RESULTS ANALYSIS

We focus on five variable groups to investigate differences in Advanced Camp performance and OML: academic tier, host tier, brigade, scholarship type, and academic discipline. We also use these groups to explore the survey responses that suggest reasons for varying Advanced Camp performance scores. In this section, we describe the flow of cadets from each grouping to Advanced Camp performance to OML.

1. Academic Tier

In this section, we present the results of our statistical analysis of academic tier in relation to Advanced Camp performance and OML. A stacked bar chart that depicts the proportions of cadets in each academic tier that achieve each Advanced Camp performance score is shown in Figure 5. As described in Chapter III, the academic tier categories are based on the Peterson's Guide for academic entrance difficulty (Haupt 2018a). We use a one-sided Jonckheere-Terpstra test to assess the significance of Advanced Camp performance differences among academic tiers because both variables are ordered categorical variables. The alternate hypothesis of the one-sided Jonckheere-Terpstra test states that there is an increasing association between academic tier and Advanced Camp performance.

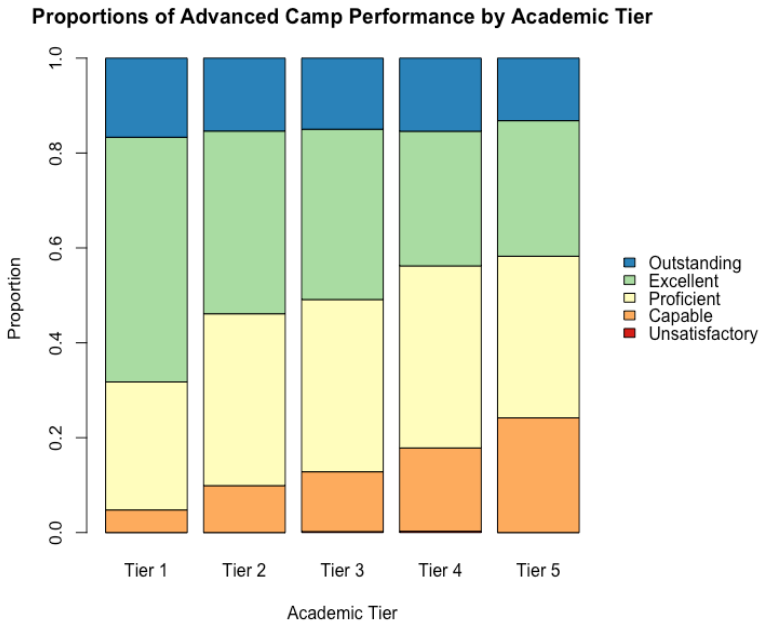

Figure 5. Proportions of Advanced Camp Performance by Academic Tier

The Jonckheere-Terpstra test produces a p-value of $1.0 \cdot 10^{-5}$ (1 in 100,000); we reject the null hypothesis that proportions of Advanced Camp performance scores are equal among Academic Tiers. Therefore, we find that as academic entrance difficulty increases, Advanced Camp performance scores improve.

A stacked bar chart that displays the proportions of cadets in each academic tier that fall into each OML quartile is shown in Figure 6. Both variables are ordered categorical variables, so we continue to use the Jonckheere-Terpstra test to assess the significance of OML quartile differences among academic tiers.

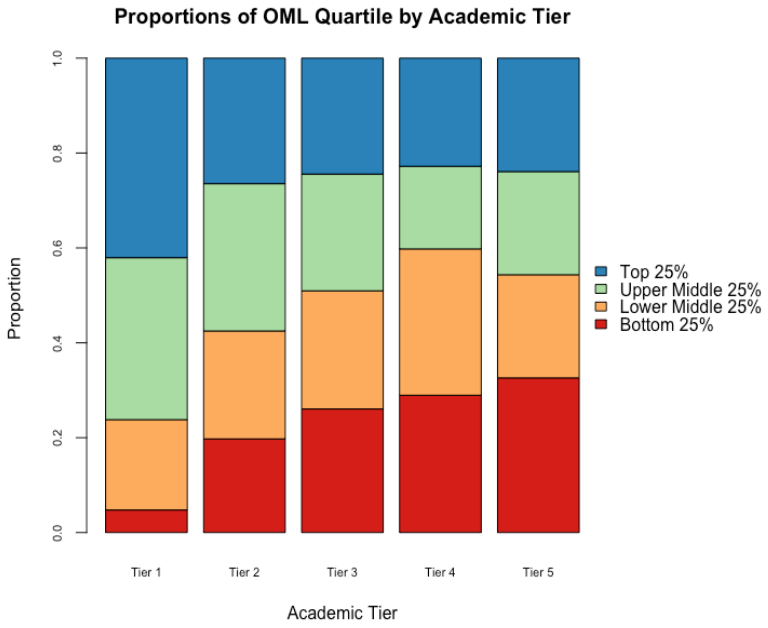

Figure 6. Proportions of OML Quartiles by Academic Tier

The Jonckheere-Terpstra test results in a p-value of 0.0007; we reject the null hypothesis that proportions of cadets in each OML quartile are equal among academic tiers. Therefore, we find that as academic entrance difficulty increases, more cadets fall into superior OML quartiles.

We use the Sankey diagram shown in Figure 7 to display the flow of cadets from academic tiers to Advanced Camp performance scores to OML quartiles. Sankey diagrams allow multiple categorical variables to be displayed simultaneously in order to display their relationships. The widths of the arrows in our Sankey diagram are proportional to the quantity of cadets in the flow. In this diagram, Tier 3 dominates the Proficient category and cadets who receive a Proficient score or lower at Advanced Camp rarely find themselves in the top 25% of OML.

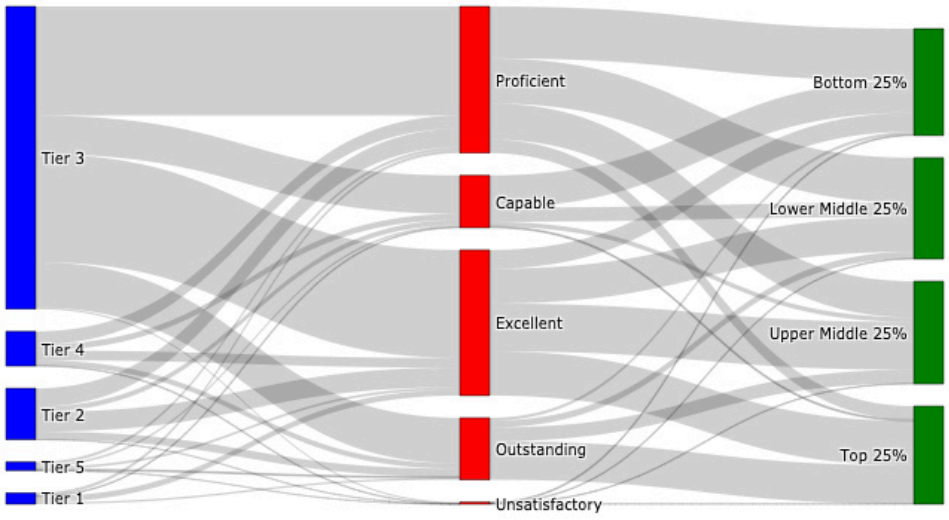

Figure 7. Flow of Cadets: Academic Tier to Advanced Camp to OML

Survey responses provide several insights about how prepared cadets are for Advanced Camp based on their academic tier. Figure 8 shows that Tier 1 cadets outperform all other tiers; however, these cadets report participating in fewer group physical training (PT) events than cadets from other tiers. This is counterintuitive because one would reasonably expect that more group PT sessions lead to better physical fitness performance. While physical fitness scores vary among tiers, there is little variation in responses to the question, "How well do you feel prepared for the Physical Fitness Test at Advanced Camp?"

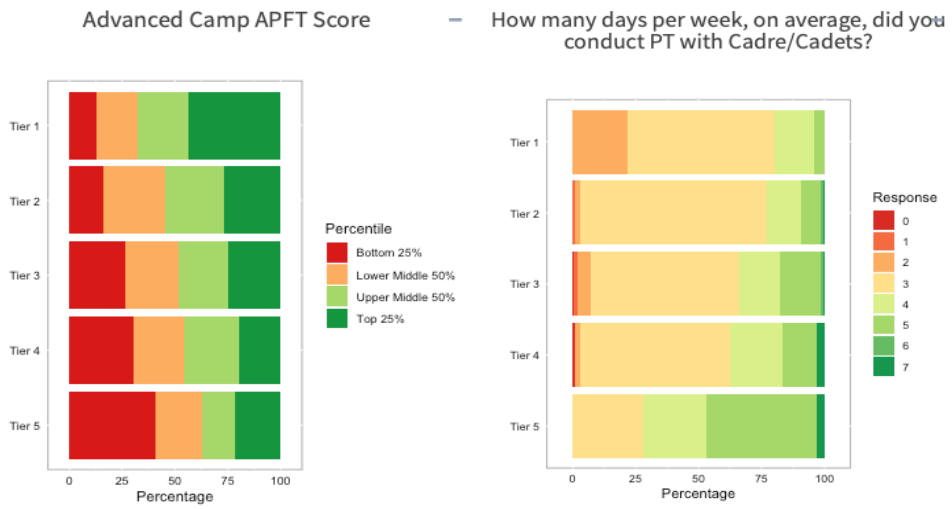

Figure 8. Survey Responses by Academic Tier: Physical Fitness

Cadets from Tier 1 outperform cadets from all other tiers on the land navigation written test. Responses to the question, "How well do you feel prepared for land navigation at Advanced Camp?" reveal that cadets across all academic tiers generally feel equally prepared for this Advanced Camp event. Land navigation written test scores and corresponding question responses are shown in Figure 9.

Figure 9. Survey Responses by Academic Tier: Land Navigation

One survey question that does yield varying responses across academic tiers is the following: "Would you describe yourself as: Always wanted to be an Army officer?" Tier 1 cadets have the smallest percentage of 'Fits Me Perfectly' responses; whereas, Tier 5 cadets have the largest percentage, as shown in Figure 10.

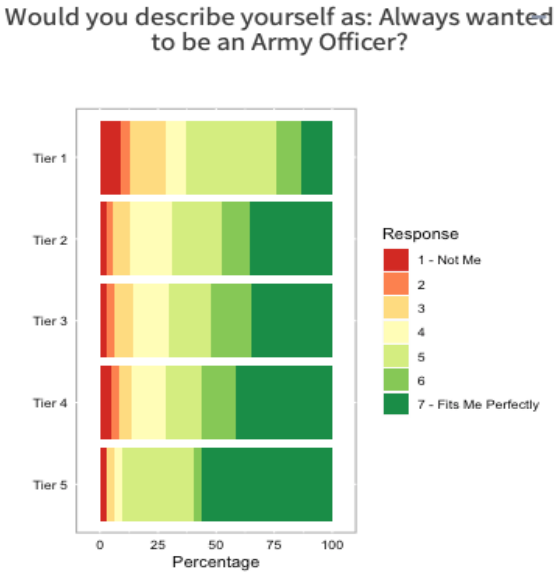

Figure 10. Survey Responses by Academic Tier: Desire to be Army Officer

2. Host Tier

We find that 4052 of 4814 cadets in our merged data set have matching academic tier and host tier values. Results from our host tier analysis mimic the results from our academic tier analysis. Appendix B contains the stacked bar charts and Sankey diagram from our host tier analysis.

3. Brigade

In this section, we discuss the results of our statistical analysis of the brigade variable in relation to Advanced Camp performance and OML. There is no order of precedence among the different brigades as they are organized geographically. Therefore, we use the Kruskal-Wallis test to assess the significance of Advanced Camp performance

and OML quartile differences among brigades. The stacked bar chart in Figure 11 shows the proportion of cadets that receive each Advanced Camp performance score, grouped by brigade.

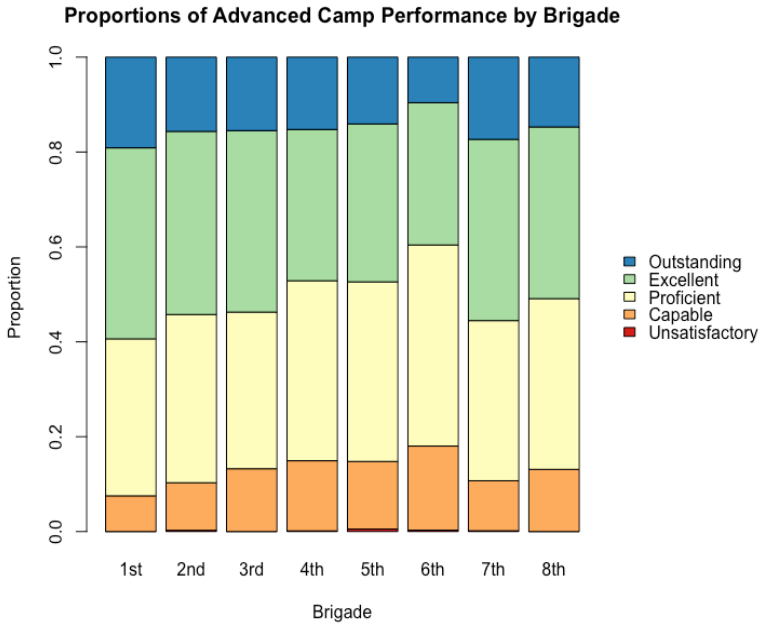

Figure 11. Proportions of Advanced Camp Performance by Brigade

The Kruskal-Wallis test gives a p-value of 0.00056, which leads us to reject the null hypothesis that brigades are from identical populations. We find that there is at least one pair of brigades with the property that one receives generally higher Advanced Camp ratings than the other. Nearly 60% of cadets from 1st Brigade receive an Outstanding or Excellent Advanced Camp performance score. Alternatively, nearly 60% of cadets from 6th Brigade receive a Proficient or Capable Advanced Camp performance score.

The stacked bar chart in Figure 12 shows the proportion of cadets that fall into each OML quartile, grouped by brigade.

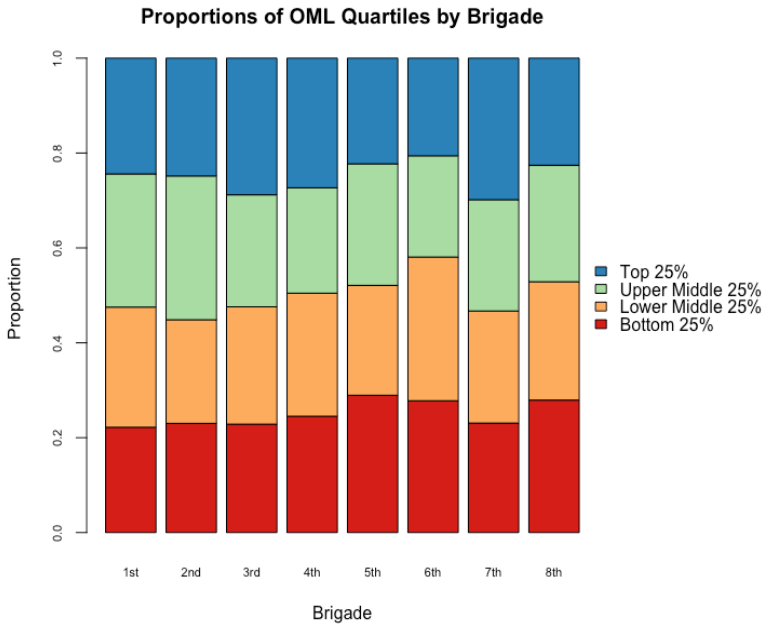

Figure 12. Proportions of OML Quartiles by Brigade

The Kruskal-Wallis test produces a p-value of 0.003, which leads us to reject the null hypothesis that the brigades come from the same population in favor of the alternate hypothesis that at least one of the brigades comes from a different population than the rest. Graphically, 7th Brigade has the largest proportion of cadets that fall in the top 25% of OML. Also, nearly 60% of 6th Brigade falls into the bottom 50% of OML.

The Sankey diagram describing the flow of cadets from brigade to Advanced Camp performance to OML is shown in Figure 13. The 2nd Brigade has the widest arrows flowing into the Excellent Advanced Camp performance score outcome, and also has a large contribution to the Outstanding outcome.

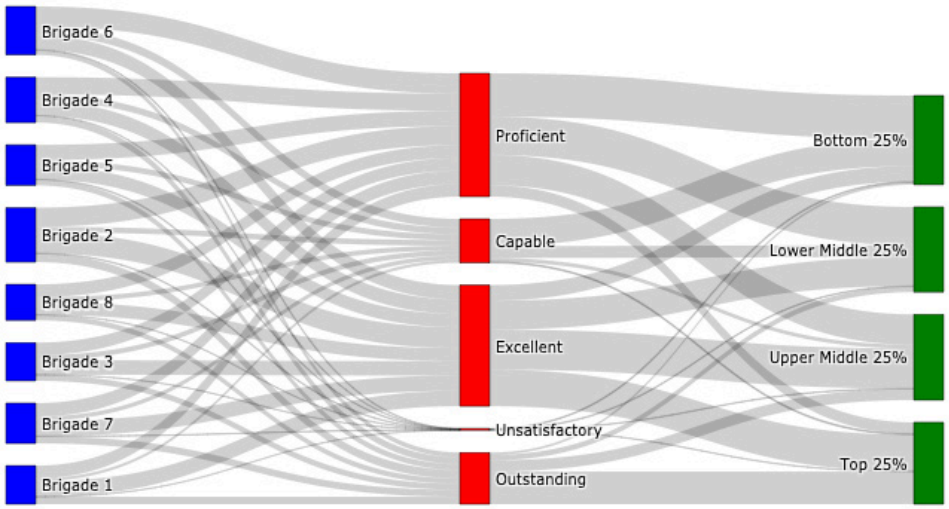

Figure 13. Flow of Cadets: Brigade to Advanced Camp to OML

Survey responses organized by Brigade generally have the same trends. There is one question where responses vary among brigades: "Would you describe yourself as: I am nervous about whether I can get through Advanced Camp?" Responses to this question are shown in Figure 14.

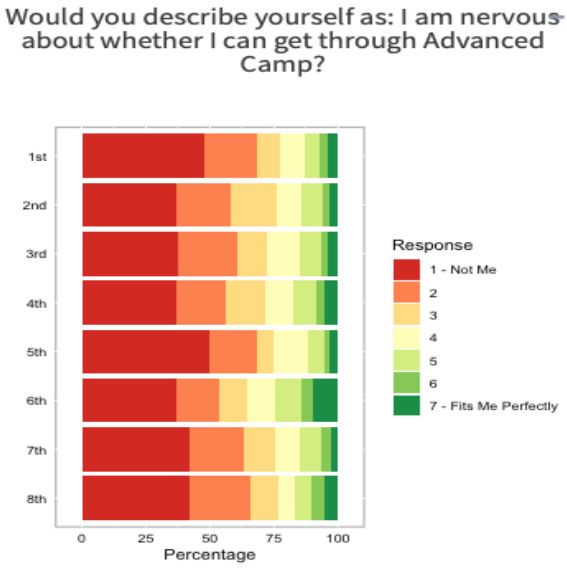

Figure 14. Survey Responses by Academic Tier: Nervous for Advanced Camp

Cadets in 6th Brigade have the largest percentage of 'Fits Me Perfectly' responses for this question. As shown previously in Figure 11, nearly 60% of cadets in 6th Brigade receive a Proficient or Capable Advanced Camp performance score. Alternatively, cadets in 5th Brigade have the largest percentage of 'Not Me' responses, yet they evenly fall into all score categories (except for Unsatisfactory) at Advanced Camp.

4. Scholarship Type

In this section, we present the results of our statistical analysis of the scholarship type variable in relation to Advanced Camp performance and OML. There is no defined hierarchy between the different types of scholarships; therefore, we use the Kruskal-Wallis test to assess the significance of Advanced Camp performance and OML differences among scholarship types. We present a stacked bar chart in Figure 15 that shows the proportions of cadets that receive each Advanced Camp performance score, grouped by scholarship type.

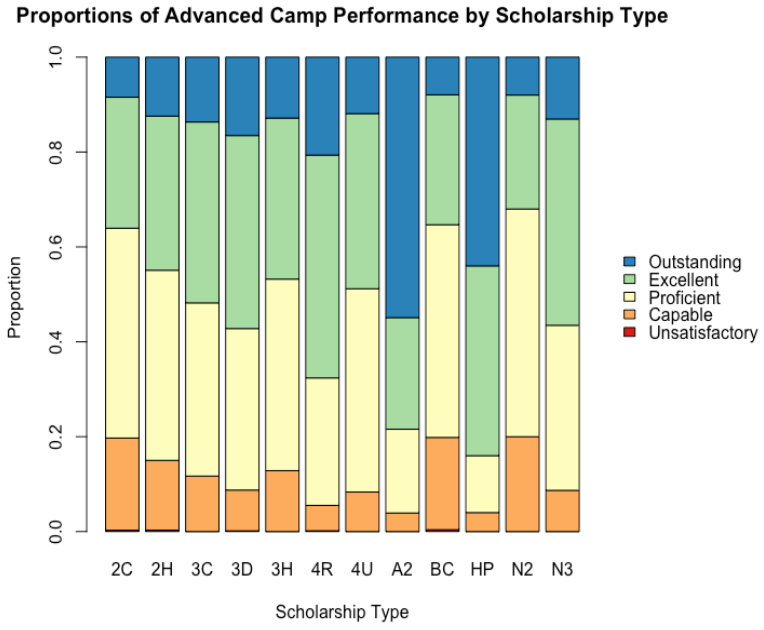

Figure 15. Proportions of Advanced Camp Performance Scores by Scholarship Type

The Kruskal-Wallis test gives a p-value of $1.788 \cdot 10^{-15}$, which leads us to reject the null hypothesis that the distribution of scholarship types all come from the same population. We find that as scholarship type changes, proportions of Advanced Camp performance scores also change. Graphically, nearly 80% of cadets with the A2 scholarship type receive an Advanced Camp performance score of either Outstanding or Excellent. Over 80% of cadets with the HP scholarship type also receive an Advanced Camp performance score of either Outstanding or Excellent. As we describe in Chapter III, the A2 scholarship represents a two-year active duty scholarship and the HP scholarship represents two-year Green-to-Gold hip pocket scholarship.

To explore the relationship between scholarship type and OML, we provide a stacked bar chart in Figure 16 that shows the proportion of cadets that fall into each OML quartile, grouped by scholarship type.

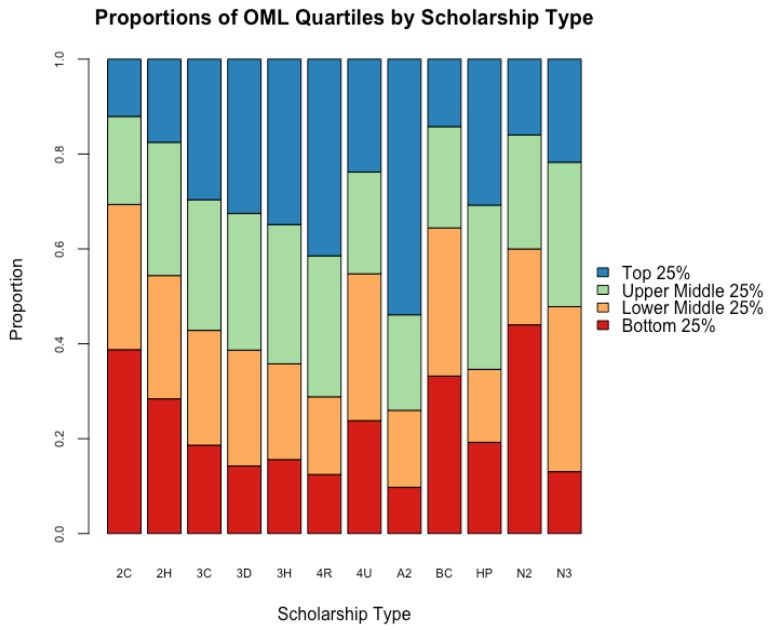

Figure 16. Proportions of OML Quartiles by Scholarship Type

The Kruskal-Wallis test results in a p-value of $2.2 \cdot 10^{-16}$, which leads us to reject the null hypothesis that our scholarship groups all come from the same population. We find

that scholarship type has bearing on a cadet's OML. In Figure 16, nearly 50% of cadets who receive the A2 scholarship fall in the top 25% of OML. Additionally, over 40% of cadets who receive the N2 scholarship type fall in the bottom 25% of OML.

The Sankey diagram displaying the flow of cadets from scholarship type to Advanced Camp performance to OML is shown in Figure 17. In this diagram, the 2H scholarship type dominates the Proficient and Excellent Advanced Camp performance scores.

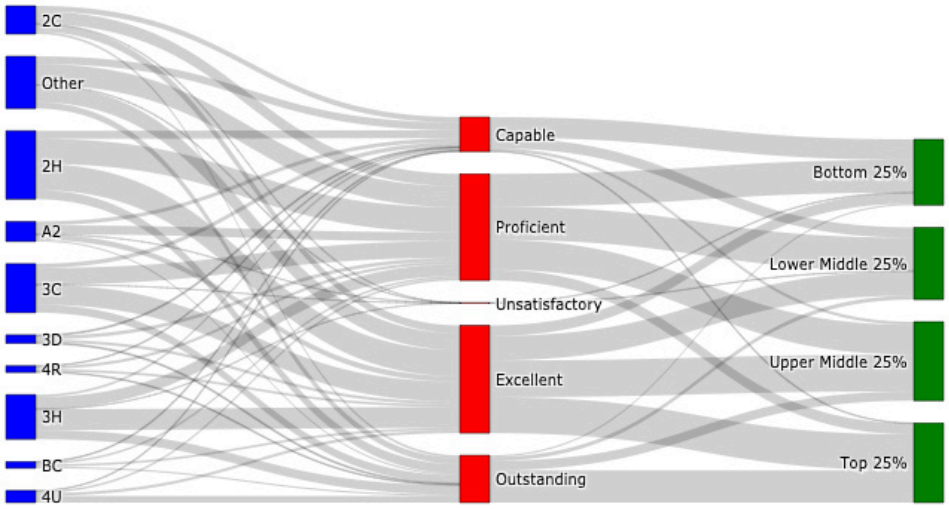

Figure 17. Flow of Cadets: Scholarship Type to Advanced Camp to OML

Several survey responses vary among cadets with different scholarship types. Figure 18 shows the Advanced Camp performance metrics for marksmanship, organized by scholarship type. Cadets who have an A2 scholarship excel at both tested marksmanship events when compared to cadets with other scholarship types.

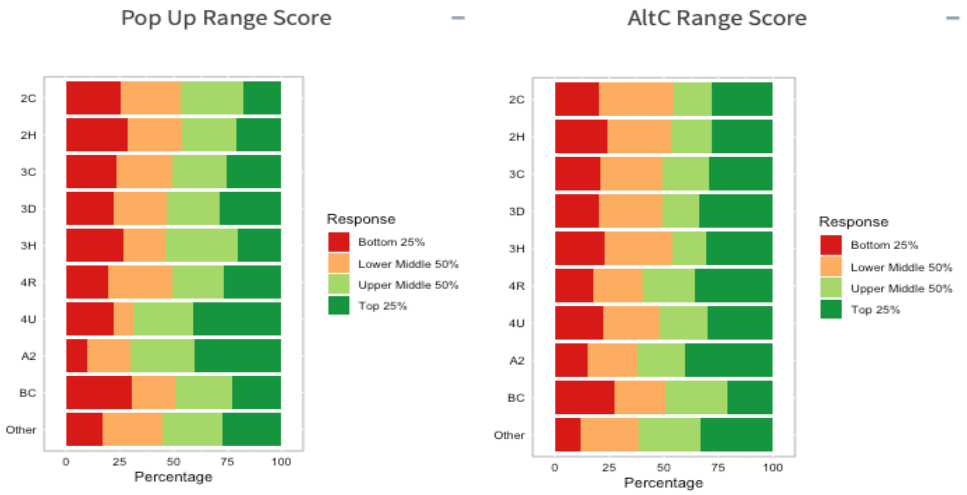

Figure 18. Advanced Camp Marksmanship Scores by Scholarship Type

Survey responses that relate to marksmanship preparation are shown in Figure 19. As described in Chapter III, A2 scholarships are two-year active duty scholarships. Cadets with A2 scholarships report having a lot of experience with rifles and feel the most prepared for basic rifle marksmanship (BRM) compared to cadets who receive other types of scholarships. We find that cadets with a lot of rifle experience and a strong feeling of being prepared directly relate to strong marksmanship scores at Advanced Camp.

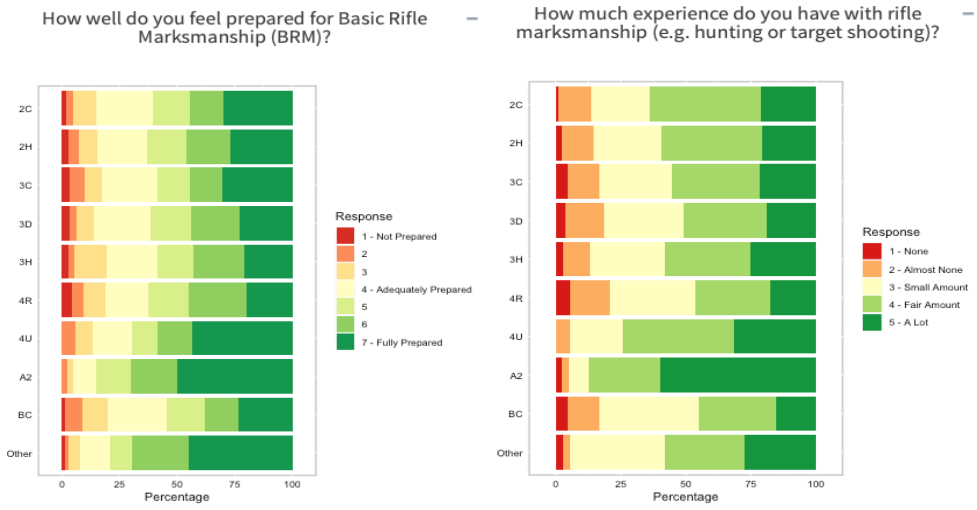

Figure 19. Survey Responses by Academic Tier: Preparation & Rifle Experience

5. Academic Discipline

In this section, we discuss our results pertaining to the relationship between academic discipline and both Advanced Camp performance and OML. For our statistical analysis, we again use the Kruskal-Wallis test because we deal with ordered categorical outcome variables grouped by an unordered categorical variable. Figure 20 shows a stacked bar chart reflecting the proportion of cadets from each academic discipline that receive each of the possible Advanced Camp performance scores.

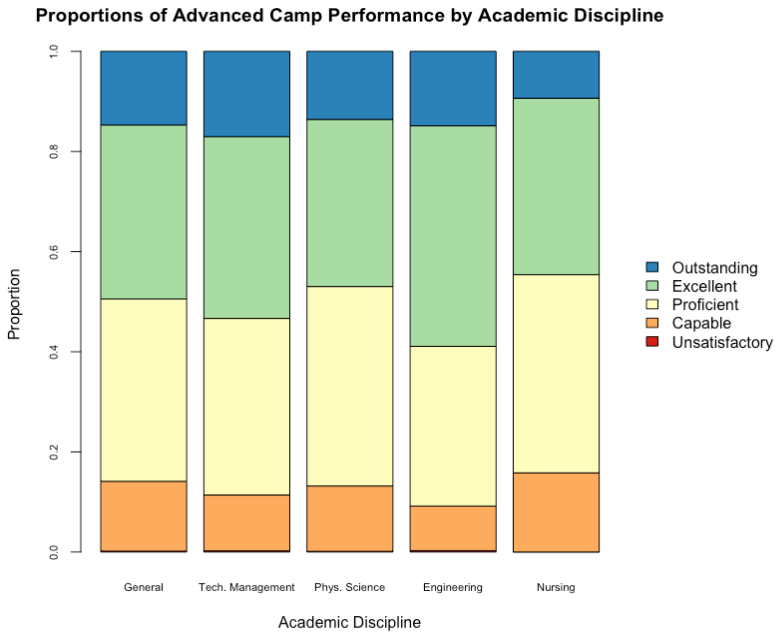

Figure 20. Proportions of Advanced Camp Performance Scores by Academic Discipline

The Kruskal-Wallis test results in a p-value of 0.2552, which causes us not to reject the null hypothesis that, in terms of Advanced Camp performance, academic discipline groups come from the same population. We find that academic discipline does not have an impact on a cadet's Advanced Camp performance score. There are no proportions in Figure 20 that graphically stand out as significantly different from the rest.

We use the stacked bar chart shown in Figure 21 to investigate the relationship between academic discipline and OML. The bar chart reflects the proportion of cadets in each academic discipline that receive each Advanced Camp performance score.

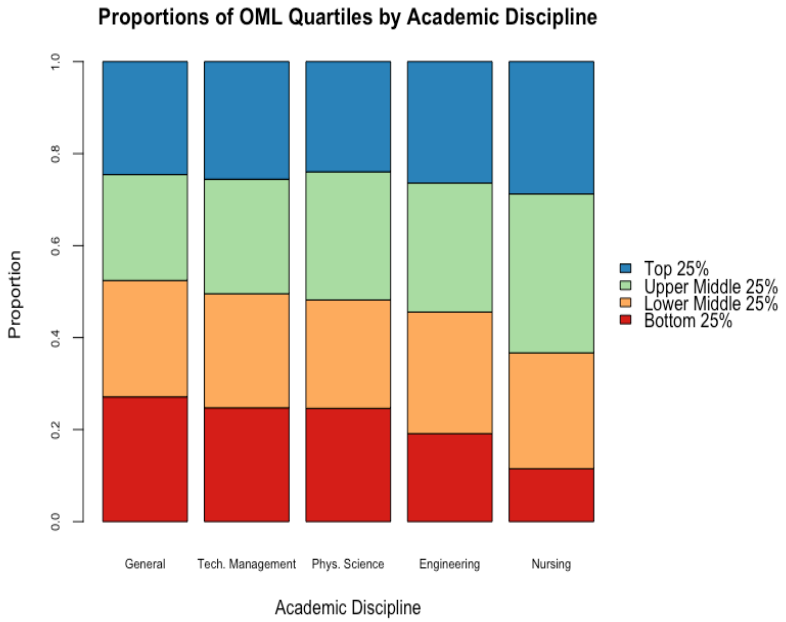

Figure 21. Proportions of OML Quartiles by Academic Discipline

The Kruskal-Wallis test gives a p-value of $4.892 \cdot 10^{-5}$; as a result, we reject the null hypothesis that, in terms of OML, academic discipline groups come from the same population. Therefore, we find that academic discipline has bearing on a cadet's OML quartile. Over 60% of cadets in the Nursing academic discipline fall into the top 50% of OML.

The flow of cadets from academic discipline to Advanced Camp performance to OML is shown using a Sankey diagram as seen in Figure 22. The General academic discipline commands the Excellent and Proficient Advanced Camp performance scores, and the Technical Management and Physical Science academic disciplines send the largest flow of cadets to the Proficient Advanced Camp performance score.

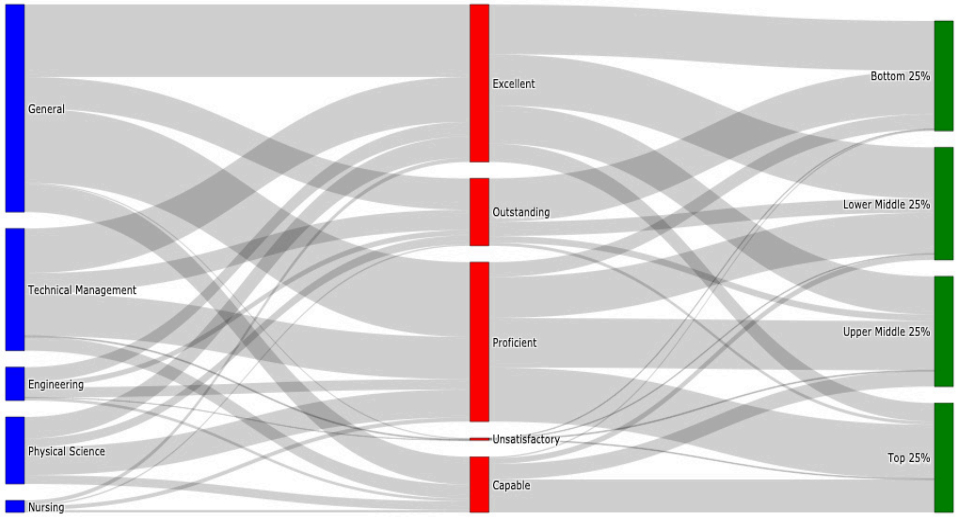

Figure 22. Flow of Cadets: Academic Discipline to Advanced Camp to OML

B. PREDICTIVE MODEL: ADVANCED CAMP PERFORMANCE

In this section, we discuss results surrounding our classification tree for predicting a cadet's performance at Advanced Camp. We use our training set to train our classification tree. As stated in Chapter III, we incorporate a loss matrix to penalize our model for misclassifications, and we focus on predictor variables that exist prior to a cadet attending Advanced Camp. We prune our classification trees using a cross-validation with one standard error threshold; we select the tree with the number of splits that has a cross-validated error under this threshold. Specifically, we include the variables shown in Table 11.

Table 11. Classification Tree Predictor Variables

Classification Tree Predictor Variables
W_MS3_OER_SCORE
W_MS3_RNK_SCORE
W_APF_FAL_RAW_SCOR
W_APF_SPR_RAW_SCOR
W_ACCESSION_GPA
W_GENDER
W_REDCAT
W_ACD_DISC_MIX
W_SCHL_CAT_CD
Brigade
Academic.Tier
Host.Tier

First, we build a classification tree without the loss matrix incorporated directly in the model. Therefore, the model maintains use of the Gini index formula shown in Equation 1 from Chapter III to maximize node purity. We only use observations that do not contain any missing values. The resulting classification tree is shown in Figure 23.

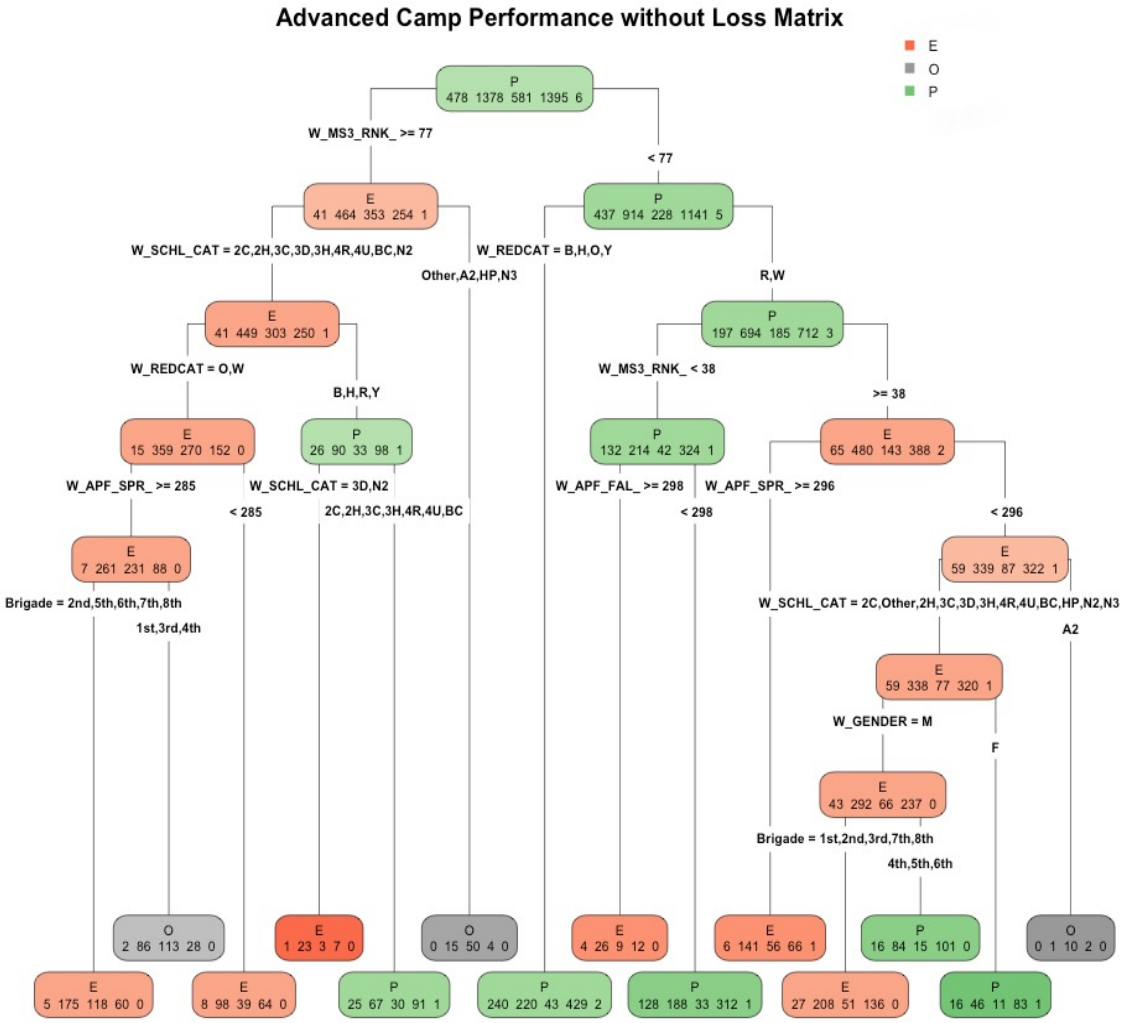

Figure 23. Classification Tree for Advanced Camp Performance without Loss Matrix

The nodes in Figure 23 are color-coded based on the dominant true classifications at each split. Nodes that are a darker shade of their respective colors are purer. Terminal nodes are located at the bottom of the tree and the letters represent the final predicted classifications of the test set observations. The five numbers listed within a node represent the frequency of each of the true classifications found in that node, in order of: C, E, O, P, U.

When we run our test set through this classification tree model, we get the resulting confusion matrix shown in Figure 24. Point-wise multiplication of this confusion matrix and our loss matrix, and then summing, gives us the total loss across all of the observations in the test set. Dividing the total loss by the number of samples in our test set gives us the average loss for a single observation. From the classification tree shown in Figure 23, and the confusion matrix shown in Figure 24, we get an average loss of 0.647. This value means that for a single prediction, we are within one degree of a cadet's actual Advanced Camp score classification.

		Actual				
		O	**E**	**P**	**C**	**U**
Predicted	**O**	37	37	18	1	0
	E	65	142	97	14	0
	P	38	156	235	117	3
	C	0	0	0	0	0
	U	0	0	0	0	0

Figure 24. Confusion Matrix for Classification Tree without Loss Matrix

Next, we build a classification tree with the loss matrix incorporated directly in the model. Now, the model uses the modified Gini index formula shown in Equation 3 from Chapter III to maximize node purity. Again, we only use observations that do not contain any missing values. The resulting classification tree is shown in Figure 25.

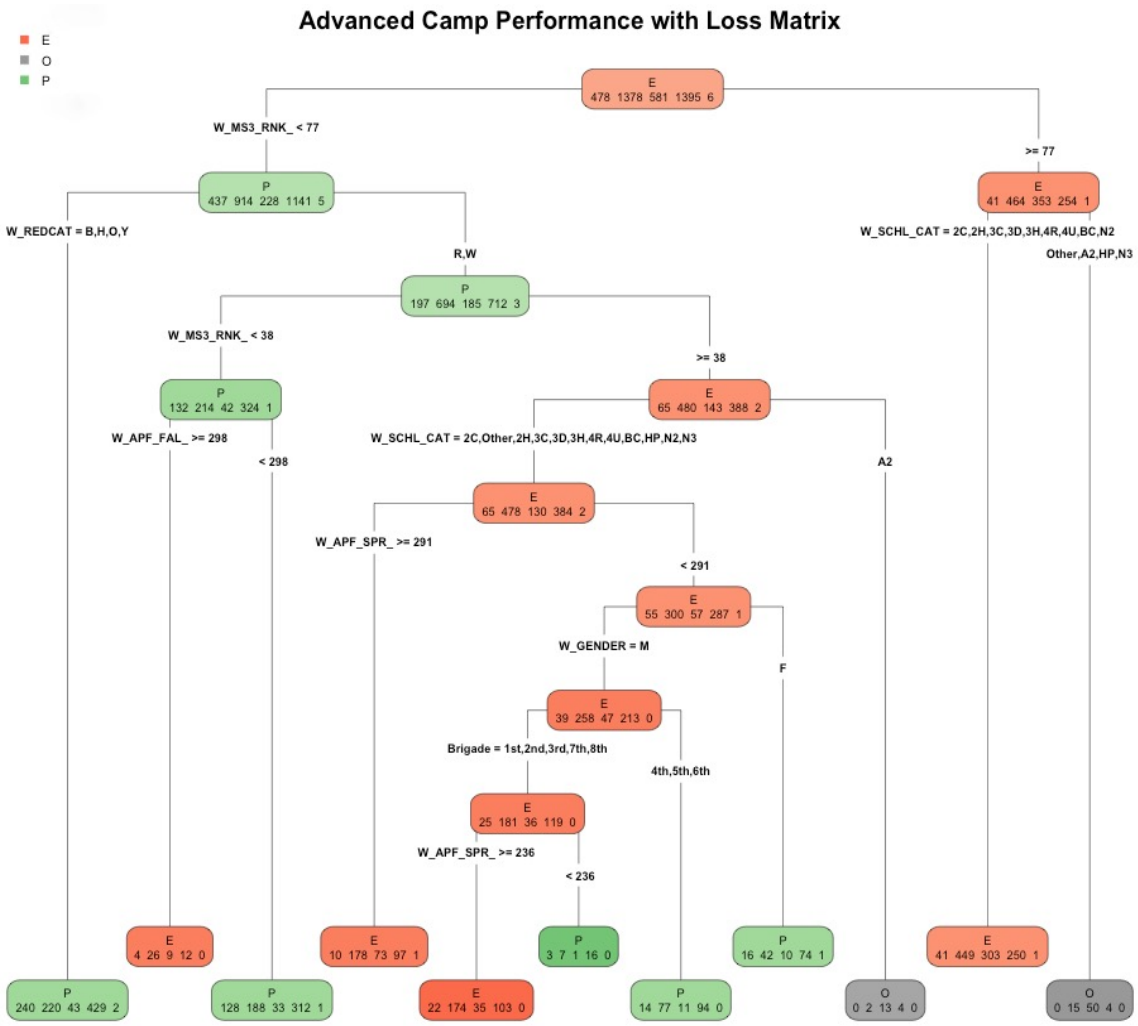

Figure 25. Classification Tree for Advanced Camp Performance with Loss Matrix

Interpretation of the classification tree shown in Figure 25 follows the same pattern described for Figure 23. Validating our tree with our test set produces the confusion matrix shown in Figure 26. Again, we multiply our confusion matrix and our loss matrix to arrive at a value that represents the total loss of our test set. We get the average loss of a single sample in our test set by dividing the total loss by the number of samples in the test set. For the tree shown in Figure 25, we arrive at an average loss of 0.622. This average loss value is an improvement upon the classification tree without the loss matrix directly incorporated into the model.

		Actual				
		O	E	P	C	U
Predicted	O	16	4	5	0	0
	E	100	187	130	23	0
	P	24	144	215	109	3
	C	0	0	0	0	0
	U	0	0	0	0	0

Figure 26. Confusion Matrix for Classification Tree with Loss Matrix

The most important predictor variable in this classification tree is W_MS3_RNK_SCORE, and it is represented by the first node split in the tree. Cadets who earn a W_MS3_RNK_SCORE ≥ 77 are likely to achieve a score of Outstanding or Excellent at Advanced Camp. Ultimately, we find our tree to be a valid predictive model for Advanced Camp Performance. An average loss of 0.622 means that we successfully predict a cadet's Advanced Camp performance score within one degree of his or her actual score. The C or U classes are never predicted because the frequency of cadets who receive these scores are lower compared to other score classifications. Because the model is operating under the penalty of a loss matrix, the model's total loss is minimized if these classifications are never predicted. This model would, at worst, predict that cadets who actually receive a score of C would get a score of P because that only has a loss of 1.

C. PREDICTIVE MODEL: OML

In this section, we present the results of our MARS model for predicting a cadet's OML. We construct two MARS models. The first model includes all of the Advanced Camp performance variables. The second model does not include any of the Advanced Camp performance variables. We compare the predictive power of the two models to investigate the importance of Advanced Camp performance for OML. While we already have the OML model used to derive the OMS scores that are then ordered to produce a cadet's rank, the OML model does not take into effect the variability and relationships of its components. Minimal variance within an OML model component may boost the importance of other components once OMS scores are ranked. We ultimately find that including Advanced Camp variables drastically improves the model's predictive power, suggesting that Advanced Camp is an integral component to OML.

1. MARS Model with Advanced Camp Variables

We first build a model to predict OML based on variables that exist up to and including Advanced Camp. The OML model described in Chapter III produces an OMS score that then gets ranked to determine a cadet's OML. It is possible that scores for some components of the OML model vary only slightly between cadets, which essentially neutralizes the value of those components. In addition to its strong predictive power for OML, our MARS model with Advanced Camp variables sheds light on the true importance of OML model components. The variables we include in our model are shown in Table 12.

Table 12. MARS Model with Advanced Camp Predictor Variables

MARS Model with Advanced Camp Predictor Variables
W_MS3_OER_SCORE
W_MS3_RNK_SCORE
W_CST_OER_SCORE
W_CST_RNK_SCORE
W_APF_FAL_RAW_SCOR
W_APF_SPR_RAW_SCOR
W_APF_CST_SCORE
W_ACCESSION_GPA
W_ACTIV_PTS_EARNED
W_TRN_PTS_EARNED
W_EACTI_PTS_EARNED
W_MAT_OMS_SCORE
W_TRN_ACT_SCORE
W_ADM_OMS_SCORE
W_ATHS_OMS_SCORE
W_CLA_SCORE

The MARS model has two parameters that can be tuned for optimal model performance: the degree of interactions between the variables and the number of terms kept in the model (Boehmke 2018). We tune our MARS model using a hyper-grid of 30 possible combinations of these parameters to arrive at our best performing model. Figure 27 shows the RMSE for the 30 possible parameter combinations in our hyper-grid. Our optimal model has a degree of three and uses 28 of 31 terms and 12 of 16 predictor variables.

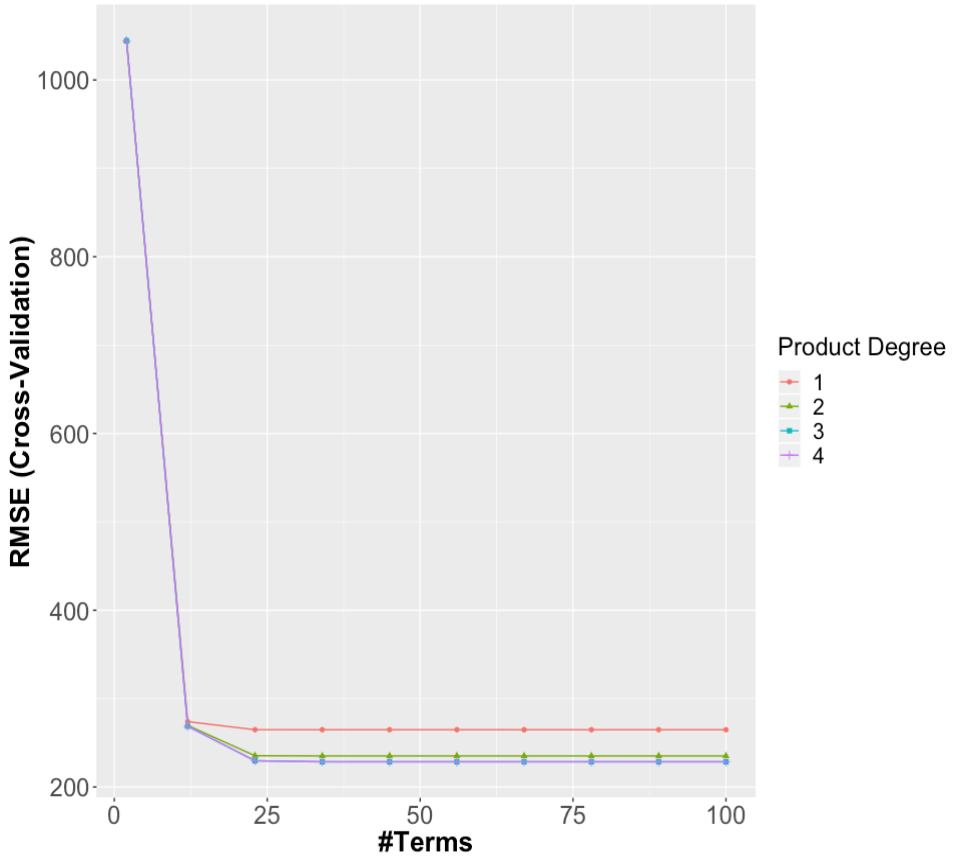

Figure 27. RMSE for Tuning MARS Model with Advanced Camp Variables

Our MARS model has an R^2 value of 0.98, demonstrating that our model accounts for approximately 98% of the variance in the data. We do not find this high R^2 value surprising given that we include all of the variables that comprise the official 2018 OML model. We explore the importance of the variables in the model through a variable importance plot based on GCV and residual sum of squares (RSS). GCV approximates true cross-validation (CV) which involves creating k folds of the training data, training the model on $k-1$ of the folds, and validating the model with the remaining fold (Faraway 2016). RSS is the sum of the squared errors of the predictions. The variable importance plot for our MARS model is shown in Figure 28.

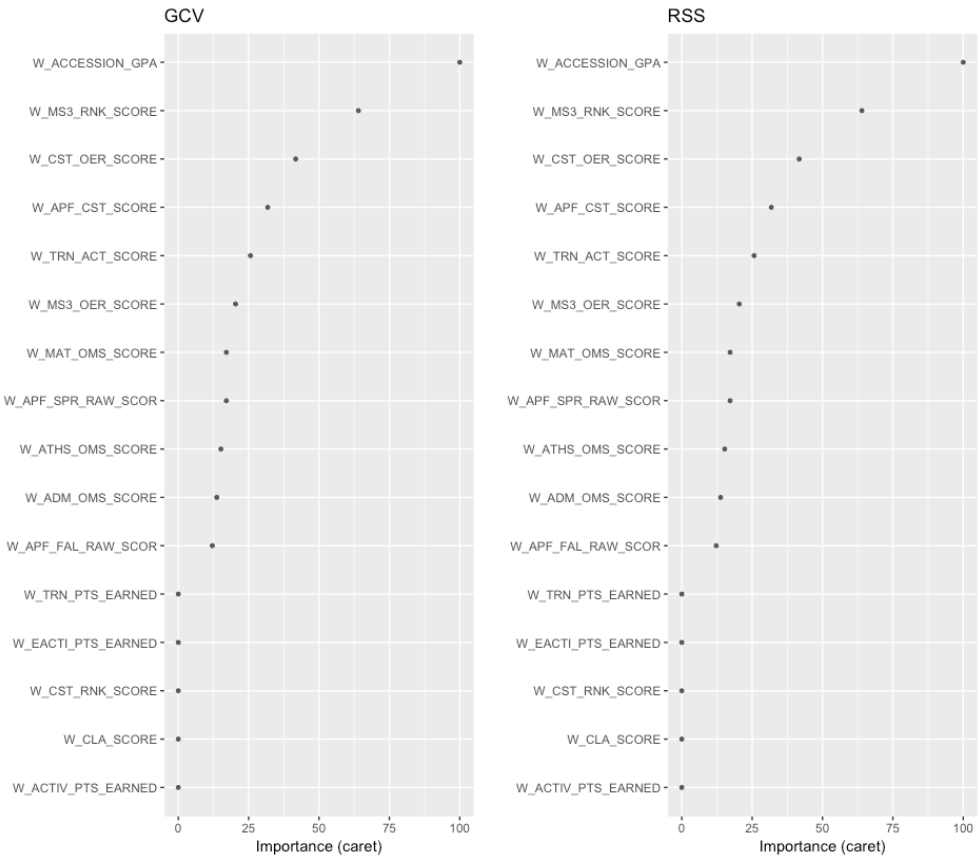

Figure 28. Variable Importance Plot for MARS Model with Advanced Camp Variables

The variable importance plots for GCV and RSS look nearly identical. The W_ACCESSION_GPA variable dominates the model. Two Advanced Camp performance variables are in the top four most important variables for the model: W_CST_OER_SCORE and W_APF_CST_SCORE. Additionally, the order of variable importance deviates from the weights of each variable seen in the OML model. Figure 29 shows the interaction between the two most important variables in this MARS model: W_ACCESSION_GPA and W_MS3_RNK_SCORE. As the values of both variables increase, a cadet's OML rank improves.

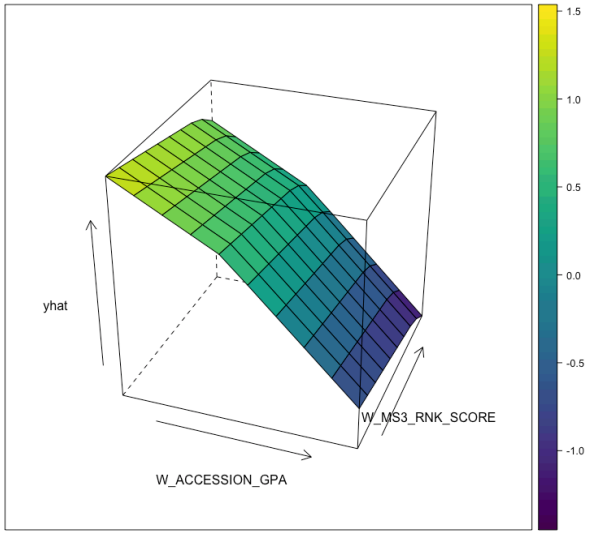

Figure 29. MARS Model with Advanced Camp Variables: Interaction Plot of Two Most Important Variables

We compare the results of our MARS model to several other regression alternatives which include multiple regression, principal component regression (PCR), partial least squares (PLS) regression, regularized regression (elastic net), and a random forest. Multiple regression is a predictive technique for multiple predictor variables as shown in Equation 4 of Chapter III. PCR uses the principal components (PCs), or linear combinations of predictors, as the predictor variables in a regression rather than the predictor variables themselves (Faraway 2014, p. 164). PLS regression is similar to PCR in that it focuses on linear combinations of the predictor variables; however, unlike PCR regression, PLS regression focuses on the response variable when determining linear combinations (Faraway 2014, p. 172). Regularized regression with Elastic Net variable selection performs automatic variable selection and is a penalized least squares method (Zhou and Hastie 2005, pp. 302–303). Finally, a random forest constructs regression trees with bootstrap samples of the data, and splits each node using the best among a subset of predictors at a given node (Liaw and Wiener 2002).

We use the R statistical programming language to build the models in our study. (R Core Team 2018). We train and tune our regression alternatives using the caret package, which stands for classification and regression training (Kuhn 2018). All of our alternatives

use a 10-fold cross-validation. The number of components included in the best tune for the PCR and PLS regression are 15 and 14, respectively. The optimal parameters for the regularized regression with Elastic Net variable selection are $\alpha = 0.7$ and $\lambda = 3.014$.

In order to put our predicted results on the same scale as the original rankings, we transform our response variable to a normal distribution using Equation 7, where Φ^{-1} is the inverse of the cumulative distribution function of a normal random variable.

$$\boldsymbol{Transformed\ OML} = \boldsymbol{\Phi^{-1}(OML\ Rank/(max\,rank + 1))} \quad (7)$$

We validate our MARS model and model alternatives using the test set with the transformed response variable. Once we have predictions, we reverse the transformation of each predicted observation value, $\hat{y}_i$, using the Equation 8.

$$\boldsymbol{\widehat{OML_i} = \Phi(\hat{y}_i) * (max\,rank + 1)} \quad (8)$$

Upon completion of the transformation, we calculate the RMSE for the MARS model and model alternatives. Table 13 shows the results of the MARS model and alternatives with both an untransformed response variable and a transformed response variable. Results shown in Table 13 are based on validating our trained models with the test set.

Table 13. RMSE for Predictive Models with Advanced Camp Variables

	Predictive Models with Advanced Camp Variables	
	RMSE: Untransformed OML Response	*RMSE: Transformed OML Response*
Multiple Regression	303.1	203.3
PCR	305.1	204.6
PLS	303.1	203.4
Elastic Net	303.0	203.1
Random Forest	274.9	257.9
MARS	227.8	192.1

Our MARS model outperforms all of the alternatives we explore in our study. We find that the MARS model predicts a cadet's OML rank within 500 out of 5538 places of his or her actual rank 98.4% of the time, within 200 places of his or her actual rank 88.5% of the time, and within 100 places of his or her actual rank 76.9% of the time. The model over-ranks and under-ranks at a nearly equal frequency and does not behave differently for subgroups. Specifically, it does not over-rank truly low individuals and does not under-rank high individuals.

2. MARS Model without Advanced Camp Variables

Next, we remove the Advanced Camp variables and repeat the same MARS modeling process. Table 14 shows the variables we use for predicting OML in the absence of Advanced Camp performance scores.

Table 14. MARS Model without Advanced Camp Predictor Variables

MARS Model without Advanced Camp Predictor Variables
W_MS3_OER_SCORE
W_MS3_RNK_SCORE
W_APF_FAL_RAW_SCOR
W_APF_SPR_RAW_SCOR
W_ACCESSION_GPA
W_ACTIV_PTS_EARNED
W_TRN_PTS_EARNED
W_EACTI_PTS_EARNED
W_MAT_OMS_SCORE
W_TRN_ACT_SCORE
W_ADM_OMS_SCORE
W_ATHS_OMS_SCORE
W_CLA_SCORE

We tune this MARS model using the same hyper-grid of 30 possible combinations of the two possible tuning parameters to arrive at our best performing model. Figure 30 shows our model's RMSE for the 30 possible parameter combinations in our hyper grid.

Our optimal model has a degree of three and uses 21 of 23 terms and 8 of 13 predictor variables.

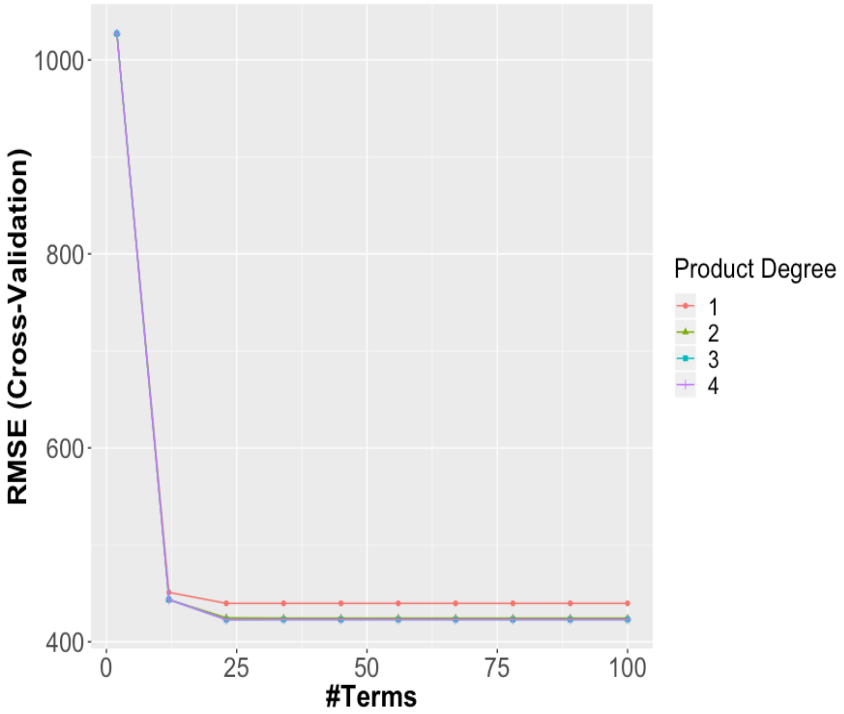

Figure 30. RMSE for Tuning MARS Model without Advanced Camp Variables

This MARS model has an R^2 value of 0.93, demonstrating that our model accounts for approximately 93% of the variance in the data. The variable importance plot for this MARS model without Advanced Camp variables is shown in Figure 31.

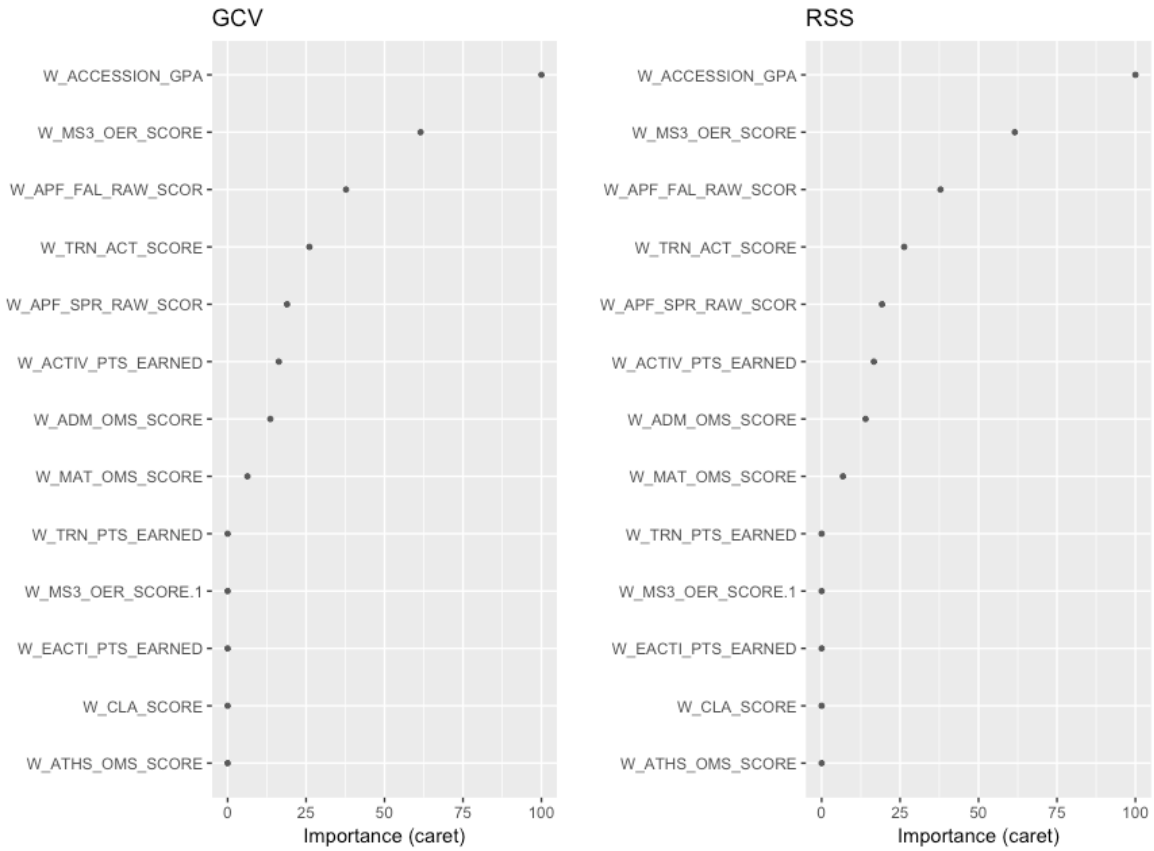

Figure 31. Variable Importance Plot for MARS Model without Advanced Camp Variables

Again, the W_ACCESSION_GPA variable dominates the model. Additionally, two of the APFT variables, W_APF_FAL_RAW_SCOR and W_APF_SPR_RAW_SCOR make up two of the top five most important variables for this model, taking the places of the absent Advanced Camp variables. The presence of two physical fitness test scores toward the top of the importance list shows that USACC and the ROTC place great value in physical abilities. Figure 32 shows the interaction between the two most important variables in this MARS model: W_ACCESSION_GPA and W_MS3_OER_SCORE. Again, as the values of both variables increase, a cadet's OML rank improves.

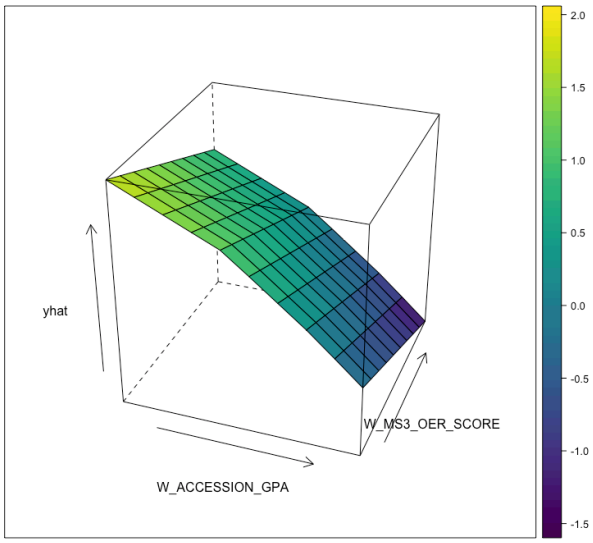

Figure 32. MARS Model without Advanced Camp Variables: Interaction Plot of Two Most Important Variables

Finally, we compare the results of our new MARS model to the alternative regression models including multiple regression, PCR, PLS regression, regularized regression (elastic net), and a random forest. Again, we perform the same transformation to our OML response variable to ensure our predicted OML values fall on the same scale of the actual OML rankings. Table 15 shows the results of the MARS model without Advanced Camp variables and model alternatives with both an untransformed response variable and a transformed response variable. Results shown in Table 15 are based on validating our trained models with the test set.

Table 15. RMSE for Predictive Models without Advanced Camp Variables

	Predictive Models without Advanced Camp Variables	
	RMSE: Untransformed OML Response	*RMSE: Transformed OML Response*
Multiple Regression	457.4	402.8
PCR	459.5	403.8
PLS	457.4	402.9
Elastic Net	457.4	402.5
Random Forest	423.7	413.0
MARS	406.5	385.9

The MARS model outperforms all of the alternatives we explore in our study. We find that this MARS model predicts a cadet's OML rank within 500 out of 5538 places of his or her actual rank 91.9% of the time, within 200 places of his or her actual rank 74.0% of the time, and within 100 places of his or her actual rank 64.0% of the time. We find that the absence of Advanced Camp performance variables in our predictive model for OML drastically reduces prediction accuracy and demonstrates the importance of strong Advanced Camp performance.

V. CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY

Our thesis examines how information about ROTC cadets prior to Advanced Camp can help us determine how well cadets are prepared for Advanced Camp. Surveys that cadets take prior to Advanced Camp provide additional insights into their preparation. Additionally, our thesis investigates whether an effective hierarchical model to predict Advanced Camp and subsequently OML is possible. We use the Jonckheere-Terpstra test for ordered differences, the Kruskal-Wallis test for differences in mean ranks among groups, and Sankey diagrams to identify subsets of cadets that show a propensity for strong or weak performance at Advanced Camp. We then use a classification tree with a loss matrix to predict Advanced Camp performance scores based on pre-existing cadet data and demographics. Finally, we tune and use a MARS model to predict a cadet's OML both with and without Advanced Camp performance variables to compare the predictive power of each model and observe the impact of a cadet's Advanced Camp performance on OML. In this section, we present the three questions we consider in our analysis with our findings.

(1) How well are students prepared for Advanced Camp? In other words, which subsets of cadets, based on cadet data and demographics, show a propensity for strong or weak performance at Advanced Camp?

We summarize our results in order of the five variable groups our study focuses on: academic tier, host tier, brigade, scholarship type, and academic discipline. For the academic tier variable, our study finds evidence that as academic difficulty increases, Advanced Camp performance improves. Over 75% of cadets enrolled in Tier 1 academic institutions fall into the top 50% of OML. Results for the host tier variable mimicked the results for the academic tier variable because nearly 85% of cadets have matching academic and host tier values. For the brigade variable, we find that nearly 60% of 1st Brigade receive an Outstanding or Excellent Advanced Camp performance score; whereas, nearly 60% of cadets from 6th Brigade receive a Proficient or Capable Advanced Camp performance score. Additionally, nearly 60% of cadets in 6th Brigade fall into the bottom 50% of OML. From our investigation of the scholarship type variable, we find that cadets with the A2

and HP scholarships outperform their peers with alternate scholarships types at both Advanced Camp and in OML. Finally, for the academic discipline variable, we find that a cadet's academic discipline does not impact his or her performance at Advanced Camp but does impact his or her OML.

(2) What insights can the surveys provide about preparation for Advanced Camp?

We summarize the results of our survey investigation using the same five variables groups: academic tier, host tier, brigade, scholarship type, and academic discipline. Overall, there was little variation in survey responses within these groups, though there are a few questions that have varying responses. For the academic tier variable, we find that even though cadets enrolled in Tier 1 academic institutions have the strongest APFT performances at Advanced Camp, they report participating in the fewest group PT events each week compared to their peers. This insight suggests that ROTC units at academic Tier 1 institutions may consider replacing group PT with other training evolutions or providing students with additional time to complete their schoolwork. Additionally, Tier 1 cadets outperform all other tiers on the land navigation written test at Advanced Camp, but survey responses about feeling prepared for land navigation at Advanced Camp are generally the same across tiers. ROTC units at non-Tier 1 institutions may consider reviewing their land navigation training procedures to ensure that they align with the grading criteria of the land navigation evaluations at Advanced Camp. We also find that over 50% of Tier 5 cadets report having always wanted to be an Army officer; whereas, only about 15% of Tier 1 cadets report the same. Again, survey responses for the host tier variable mimic the survey responses for the academic tier variable.

From our investigation of the brigade variable, we find one question with responses that vary between the brigades: "Would you describe yourself as: I am nervous about whether I can get through Advanced Camp?" Cadets in 6th Brigade report the highest percentage of responses near to and including the 'Fits Me Perfectly' response, and nearly 60% of cadets in 6th Brigade receive a Proficient or Capable Advanced Camp performance score. Cadets in 5th Brigade report the highest percentage of 'Not Me' responses yet they evenly fall into all score categories (except for Unsatisfactory) at Advanced Camp. These

responses suggest that ROTC unit leadership should ensure that they provide continuous feedback to cadets throughout their junior year regarding their performance in evolutions that will be graded at Advanced Camp and tailor training events accordingly.

We find that survey response by scholarship type vary the most in the marksmanship category. Cadets who receive the A2 scholarship outperform their peers on both graded marksmanship events at Advanced Camp; these same cadets report having the most rifle marksmanship experience and the strongest feeling of preparation for Advanced Camp compared to cadets who receive other scholarship types. Based on these survey responses, rifle experience is crucial to performance at Advanced Camp, and ROTC units should incorporate more marksmanship training into their programs.

(3) Can a hierarchical model effectively predict overall Advanced Camp performance scores and subsequently OML? How important is Advanced Camp for determining OML?

We provide evidence that a classification tree with a loss matrix can predict a cadet's Advanced Camp performance score within one degree of his or her actual score using strictly pre-existing cadet data and demographic information. We find the most significant variable in this model to be the W_MS3_RNK_SCORE variable, as this is the first split that occurs in the classification tree. As a result of the low frequency of cadets who receive an Unsatisfactory Advanced Camp performance score and the incorporation of a penalizing loss matrix, we find that our tree does not predict Capable or Unsatisfactory performance scores.

Our MARS model with Advanced Camp variables successfully predicts a cadet's OML rank within 500 places of his or her actual rank 98.4% of the time, within 200 places of his or her actual rank 88.5% of the time, and within 100 places of his or her actual rank 76.9% of the time. When we remove the Advanced Camp variables from the MARS model, these percentages changes such that abbreviated model predicts a cadet's OML rank within 500 places of his or her actual rank 91.9% of the time, within 200 places of his or her actual rank 74.0% of the time, and within 100 places of his or her actual rank 64.0% of the time. The reduced accuracy in our abbreviated model suggests that Advanced Camp is a vital component of OML, perhaps more than the 16% indicated in the OML model leads on.

B. CONCLUSIONS

Our research highlights several subsets of cadets that show a propensity for strong performance at Advanced Camp and also find themselves in the upper echelons of OML. With the belief that improving cadet performance throughout the ROTC program can enhance the U.S. Army officer corps, we identify areas where improvements may be made. Specifically, investing in more cadets at academically rigorous institutions (Tier 1 and Tier 2) may elevate ROTC performance both at Advanced Camp and in overall OML. Additionally, re-allocating scholarships via increasing the number of two-year active duty scholarships and Green-to-Gold hip pocket scholarships may also enhance cadet performance throughout the ROTC program.

Our research also highlights the importance of Advanced Camp variables for determining OML. We produce a model that can predict a cadet's Advanced Camp performance score within one degree of his or her actual score. Additionally, we produce a model that can accurately predict a cadet's OML with all pre-existing cadet data up to and including Advanced Camp performance scores. Ultimately, predictions from our model can help identify OML model components that, if improved, would provide a cadet the best opportunity to achieve the greatest leap in his or her OML rank. Additionally, understanding the true importance of Advanced Camp performance for determining OML in FY19 can influence the development of future OML models.

C. RECOMMENDATIONS FOR FUTURE WORK

Based on conclusions from our research, we suggest the following two areas as future work.

First, we suggest collecting data from ROTC units directly, vice from cadets, about the amount and type of training they conduct each semester. If units have a means to directly log training events, it would offset any bias or forgetfulness on behalf of nervous cadets about to begin a stressful training event. Additionally, we recommend reviewing the survey instrument to ensure that all areas of interest are captured within the questions and that any questions that do not contribute valuable information are removed. Many questions

in the survey about cadets' feelings toward preparation for Advanced Camp show little variation in the responses.

Second, we recommend a longitudinal study that tracks ROTC-produced Army officers throughout their initial active duty commitments to observe the correlation between a cadet's Advanced Camp performance score and his or her corresponding performance as an officer. Understanding the true relationship between Advanced Camp and officer performance will allow for an improved weighting in the OML model when ranking cadets for their future branch assignments.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. ADVANCED CAMP SURVEY INSTRUMENT

PRE

2018 Alternate Pre-Advanced Training Assessment

The purpose of this training assessment is to gather Cadet feedback on your preparation for Advanced Camp (AC). The results of this training assessment will be used to assess the MS III Curriculum and make recommendations for improvement. Your information will be maintained carefully and the responses you provide will be aggregated with fellow Cadets for reporting. No individual responses will be disseminated. Please be frank with your responses.

CST Unit. What is your CST Unit?
Regiment _____
Company _____
Platoon _____

Enter the first letter of your last name followed by the last 4 of your SSN. This information will be used to link your pre- and post-camp assessment feedback, then be removed from the analysis data sets.

Next, enter a valid email address that you can access from your mobile phone. This will be used to send you the post-camp assessment link.
First Letter and Last 4: For example: D9403 _____
Email address: _____

What was your experience like in traveling to Fort Knox?

	No Problems	Some Problems	Serious Problems
Flights / transportation	1 O	2 O	3 O
In-processing	1 O	2 O	3 O
Have all paperwork from home program	O	O	O
Pay Process	O	O	O

Physical Readiness Preparation:
What was your initial APFT score at the beginning of MS III year? _____
What was your final APFT score at the end of MS III year? _____
How many days per week, on average, did you conduct PT with Cadre/Cadets? _____
How many days per week, on average, did you conduct PT on your own? _____
How many 4 mile foot marches did you conduct in your MS III year? _____
How many 6 mile foot marches did you conduct in your MS III year? _____
How many 12 mile foot marches did you conduct in your MS III year? _____
What was your best 12 mile foot march time during MS III year in hours and minutes (e.g. 3:25) _____

Land Navigation Preparation:
How many Day Land Nav courses did you conduct during your MS III year? _____
How many Night Land Nav courses did you conduct during your MS III year? _____
What was your final map reading score (as a percentage) at the end of MS III year? _____

Leadership and Counseling Preparation:

	1 Yes	2 No
Did you receive 4 or more evaluated leadership positions during your MS III year?	O	O
Were at least 2 of your MS III evaluations during an FTX?	O	O
Were you provided attributes/competency areas to improve during fall semester of MS III year?	O	O
Were you provided attributes/competency areas to improve during spring semester of MS III year?	O	O
Did you receive an initial counseling at the beginning of your MS III year?	O	O
Did you receive a final counseling at the end of your MS III year?	O	O

0 = did not answer

Other Preparation:

	1 Yes	2 No	3 I don't know
Did you conduct CBRN training (don mask, basic decon) during your MS III year?	O	O	O
Did you achieve the CBRN training standard for Advanced Camp?	O	O	O
Did you conduct Tactical Combined Casualty Care (TC3) training during your MS III year?	O	O	O
Did you achieve the TC3 standard for Advanced Camp?	O	O	O
Did you conduct Call For Fire (CFF) training during your MS III year?	O	O	O
Did you achieve the CFF standard for Advanced Camp?	O	O	O
Did you successfully rappel during your MS III year?	O	O	O
Did you conduct at least one confidence course during your MS III year?	O	O	O
Did you conduct Basic Rifle Marksmanship (BRM) training during your MS III year?	O	O	O
Did you achieve the BRM standard for Advance Camp?	O	O	O

On a scale from 1 to 7, of 1 meaning "Not At All" to 7 meaning 'Completely', please rate the following areas:

	Not At All (1)	2	3	4	5	6	Completely (7)
Did your cadre cover Atropian news and scenario topics during your MS-III year?	O	O	O	O	O	O	O
How well do you understand what will be assessed during your training?	O	O	O	O	O	O	O
Do you expect rigorous training at Advanced Camp?	O	O	O	O	O	O	O
Do you think the training on campus has prepared you for the experience you will encounter at Advanced Camp?	O	O	O	O	O	O	O

On a scale from 1 to 7, with 1 meaning "Not Me" to 7 meaning "Fits Me Perfectly", would you describe yourself as:

	Not Me (1)	2	3	4	5	6	Fits Me Perfectly (7)
Very physically fit	O	O	O	O	O	O	O
Very academically accomplished	O	O	O	O	O	O	O
A natural leader	O	O	O	O	O	O	O
Pay close attention to details	O	O	O	O	O	O	O
Have the trust of my fellow Cadets	O	O	O	O	O	O	O
Always wanted to be an Army officer	O	O	O	O	O	O	O
Parent(s) support my decision to commission	O	O	O	O	O	O	O
Not sure that I want to commission	O	O	O	O	O	O	O
I am nervous about whether I can get through Advanced Camp	O	O	O	O	O	O	O
I expect this to be a low-stress training program	O	O	O	O	O	O	O
I tend to think outside the box	O	O	O	O	O	O	O
I am easily distracted	O	O	O	O	O	O	O
I love challenges	O	O	O	O	O	O	O
I do not like doing more than one thing at a time	O	O	O	O	O	O	O

How is ROTC offered at your institution?

- O My school is the host
- O I travel to the host school for classes
- O Cadre from the host are stationed at my school
- O Cadre from the host travel to my school just for ROTC classes

When did you first decide to enroll in Army ROTC?

- O Before High School
- O Freshman or Sophomore in High School
- O Junior in High School
- O Senior in High School
- O Between HS and College
- O Freshman in College
- O College Sophomore or Later
- O After Enlisting (Prior Service)

Which of your family members are in the military or have served in the military (including step family)?

❑ Brother	❑ Mother	❑ Cousin or other relative
❑ Sister	❑ Spouse	❑ Grandparent
❑ Father	❑ Uncle / Aunt	❑ None

In this section, on a scale from 1 - 7 with 1 meaning "Not At All Prepared" to 7 meaning "Fully Prepared, rate how well you feel prepared for each training event at Advanced Camp:

	Not At All Prepared (1)	2	3	Adequately Prepared (4)	5	6	Fully Prepared (7)
Field Leader Reaction Course	O	O	O	O	O	O	O
Tactical Combined Casualty Care (TC3)	O	O	O	O	O	O	O
Patrolling and Tactical Operations	O	O	O	O	O	O	O
Cross Cultural Competence	O	O	O	O	O	O	O
Call for Fire	O	O	O	O	O	O	O
Basic Rifle Marksmanship (BRM)	O	O	O	O	O	O	O
Land Navigation	O	O	O	O	O	O	O
Physical Fitness Test	O	O	O	O	O	O	O
Troop Leading Procedures	O	O	O	O	O	O	O

In what year did you attend Basic Camp or CIET?

O I did not attend Basic Camp or CIET
O 2017
O 2016
O 2015

	None	Almost None	Small Amount	Fair Amount	A Lot
How much experience do you have with rifle marksmanship (e.g. hunting or target shooting)?	O	O	O	O	O
Please indicate your level of involvement in sports over the last 5 years	O	O	O	O	O
How much knowledge do you have of sport or performance psychology techniques to increase emotional control and mental focus under pressure?	O	O	O	O	O

Your survey responses and scores during Advanced Camp will be used to evaluate and improve training. In the future, researchers may want to use this data for other research purposes. If this occurs, the data would not include your name or other personally identifying information. Do you give consent to your data being used anonymously in future research?

O Yes
O No

Figure 33. Advanced Camp Survey Instrument. Source: U.S. Army Cadet Command (2018a).

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. PLOTS FOR HOST TIER VARIABLE

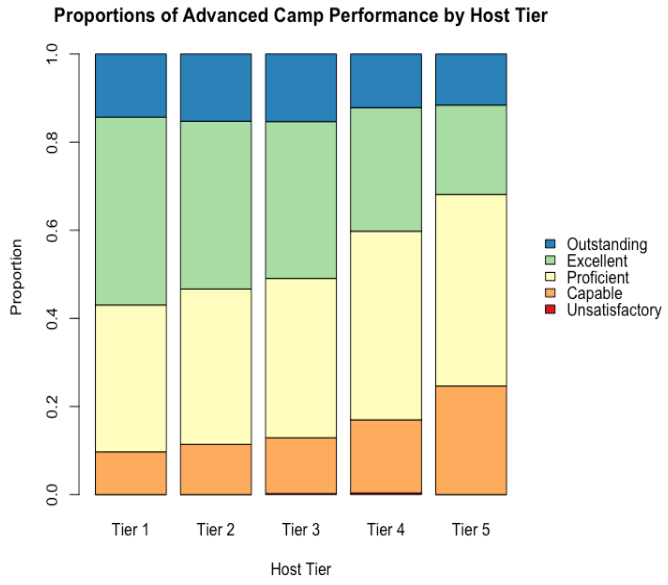

Figure 34. Proportions of Advanced Camp Performance by Host Tier

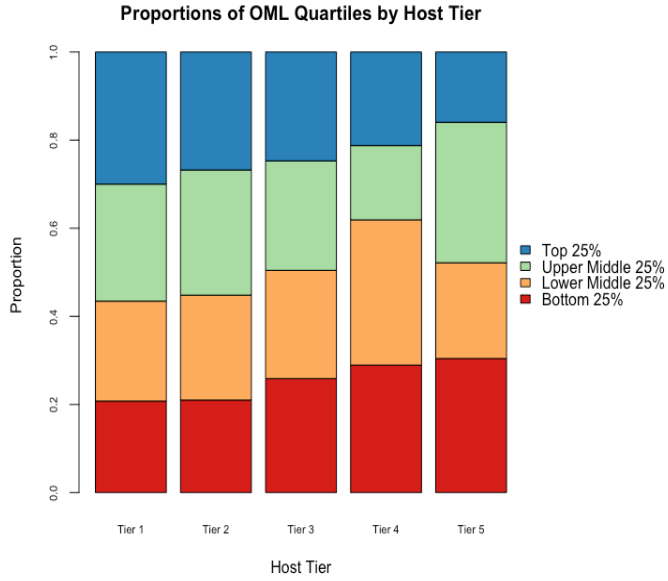

Figure 35. Proportions of OML Quartiles by Host Tier

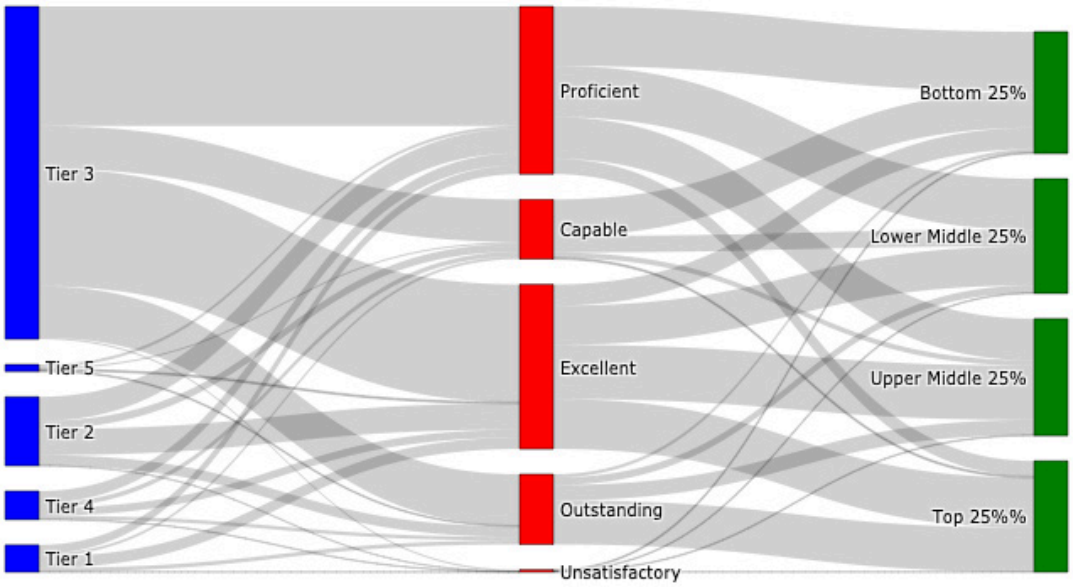

Figure 36. Flow of Cadets: Host Tier to Advanced Camp to OML

LIST OF REFERENCES

Boehmke B (2018) Multivariate adaptive regression splines. Programming guide, University of Cincinnati, http://uc-r.github.io/mars.

Department of the Army (2012a) Army physical readiness training. FM-7-22. Washington, DC, https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN7938_FM%207-22%20INC%20C1%20Final.pdf.

Department of the Army (2012b) Army leadership. ADP 6–22. Washington, DC, http://data.cape.army.mil/web/repository/doctrine/adp6-22.pdf.

Devore JL (2016) *Probability and Statistics for Engineering and the Sciences*, 9th ed. (Cengage Learning).

Duncombe FJ (2018) Policy Memorandum 9 - Advanced Camp Evaluation Report (ACER) Guidelines provided to the author by USACC, June 4.

Faraway JJ (2014) *Linear Models with R,* 2nd ed. (CRC, Boca Raton, FL).

Faraway JJ (2016) *Extending the Linear Model with R*, 2nd ed. (CRC, Boca Raton, FL).

Haupt A (2018a) Academic tier category definitions provided to the author via personal communication, October 23.

Haupt A (2018b) USACC intro to NPS. Presentation slides, October 11. U.S. Army Cadet Command.

Hastie T, Tibshirani R, Friedman J (2017) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* 2nd ed. (Springer, New York).

Hughes C (2017) Tougher summer training builds better ROTC cadets. *Army* 67(7), https://search.proquest.com/docview/1917337749/fulltextPDF/523E8B8FFFAE4DBAPQ/1?accountid=12702.

Kloke J, McKean JW (2014) *Nonparametric Statistical Methods Using R* (Chapman and Hall/CRC).

Kuhn M (2018) Caret: classification and regression training, R package version 6.0-81. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/package=caret.

Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* 2(3), https://www.researchgate.net/profile/Andy_Liaw/publication/228451484_Classification.

McCaleb BE (2016) Identifying U.S. Marine Corps recruit characteristics that correspond to success in specific occupational fields. Master's thesis, Operations Research, Naval Postgraduate School, Monterey, CA. https://calhoun.nps.edu/bitstream/handle/10945/49342/16Jun_McCaleb_Ben.pdf?sequence=1&isAllowed=y.

Office of the Under Secretary of Defense, Personnel, and Readiness (2017) Table B-32. Active component commissioned officer gains, FY17: By source of commission, service, and race/ethnicity. Accessed January 30, 2018, https://www.cna.org/pop-rep/2017/appendixb/b_32.html.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Therneau TM, Atkinson EJ (2018) An introduction to recursive partitioning using the RPART routines. Technical report, Mayo Foundation, Rochester, MN. https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

U.S. Army Cadet Command (2017a) Advanced camp. Accessed December 17, 2018, http://www.cadetcommand.army.mil/advanced.aspx.

U.S. Army Cadet Command (2017b) Brigades. Accessed December 17, 2018, http://www.cadetcommand.army.mil/brigades.aspx.

U.S. Army Cadet Command (2017c) History. Accessed December 17, 2018, http://www.cadetcommand.army.mil/history.aspx.

U.S. Army Cadet Command (2018a) Data provided to the author via TRAC Monterey, October 23.

U.S. Army Cadet Command (2018b) Reserve Officer's Training Corps accessions fiscal year 2019. USACC Circular 601–19-1. Fort Knox, KY, http://www.cadetcommand.army.mil/res/ files/forms_policies/circulars/ USACC%20Circular%20601-19-1%20Reserve%20Officers%27%20 Training%20Corps%%20 Corps% 20Accessions%20Fiscal%20Year%202018%2006-21-2018.pdf.

Zhou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J. R. Statist. Soc: Series B.* 67(2). http://www.recognition.mccme.ru/pub/papers/L1/elasticnet.pdf.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California